NAVAL MEDICAL RESEARCH UNIT – DAYTON

Development of Scoring Procedures for the Performance Based Measurement (PBM) Test: Psychometric and Criterion Validity Investigation

J. B. Phillips, O. S. Chernyshenko, S. Stark,
F. Drasgow, & H. L. Phillips, IV

NAMRU-D Report Number 12-10

Reviewed and Approved
29 November 2011

Keith A. Syring, CAPT, MSC, USN
Commanding Officer

The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government.

This work was funded by work unit number 70804.

The study protocol was approved by the Naval Aerospace Medical Research Laboratory Institutional Review Board in compliance with all applicable Federal regulations governing the protection of human subjects.

EXECUTIVE SUMMARY

This report describes findings regarding the scoring of the PBM Test and the relationships of various classical test theory (CTT) and item response theory (IRT) based subtest scores and composites with performance criteria for Navy and Marine Corps student pilots and flight officers. Overall, the IRT analyses indicated that the three parameter logistic model (3PLM) and Samejima's graded response model (SGRM) provided good fit to dichotomously and polytomously scored item-level data, respectively, for six of the seven PBM subtests. (The Dichotic Listening Test data could not be examined using IRT for reasons described within this report.) These analyses set the stage for future research involving differential item and test functioning. It was also found that PBM component scores based on examinee responses, reaction time, or tracking information often yielded criterion related validities in the range of .20 - .35. Most of the PBM component scores did not correlate highly with scores on the Aviation Selection Test Battery (ASTB); therefore, PBM subtests significantly added incremental validity in predicting training grades. Increased in incremental validity resulting from the addition of the PBM composite was highest for training blocks with higher ecological validity to actual in-cockpit performance.

In this analysis, IRT-based scores did not outperform CTT-based scores, therefore we recommend using subtest level CTT-based multiple regression composites for operational decision making. The most predictive composite that emerged in this analysis was based on six PBM sub-scores in addition to the Pilot Flight Aptitude Ratings generated by the ASTB. The CTT based PBM composite that best predicted primary flight school performance was composed of the following PBM sub-scores: Airplane Tracking Task Average Distance Z-score, the Vertical Tracking Task Average Distance Z-score, the Directional Orientation Test Total Correct, the Directional Orientation Test Total Time, the Multi Tracking Test Dichotic Listening Tests Total Correct, and the Emergency Scenarios Test Scenario Score. These results suggest that the addition of the PBM to ASTB for Naval Aviation selection will significantly reduce attrition from the Naval Aviation training pipeline, thereby saving Naval Aviation millions in training costs. Because our analysis did not include a hold-out sample, the predictive validity of this composite score should be confirmed using a new sample of Naval Aviation students.

The following sections of this report describe the PBM subtests, the demographics of the examinee samples, and the results of the psychometric and statistical analyses showing that the PBM subtests are valid predictors of a wide variety of training criteria.

Table of Contents

Page

List of Tables

List of Tables (cont.)

List of Tables (cont.)

List of Tables (cont.)

List of Tables (cont.)

List of Figures

PBM OVERVIEW

PBM Description

The Performance Based Measurement (PBM) Test is an interactive, performance-based test that is being examined for inclusion in the US Navy's web-based APEX.NET Aviation Selection Test Battery (ASTB). The test is aimed at expanding the Navy's aviation selection capabilities beyond the knowledge, skills, and abilities tapped by the existing selection battery. PBM focuses on the assessment of skills and abilities relevant to performance in flight such as audio information processing, spatial orientation, physical dexterity, divided attention, task prioritization, and decision-making.

The Services have a long history of test development for the purpose of pilot selection (e.g., Melton, 1947). Imhoff and Levine (1981) conducted a comprehensive review of the perceptual-motor skills and cognitive processes potentially related to pilot training and performance. They focused on aspects of performance that might be best assessed via hands-on tasks, rather than paper-and-pencil multiple choice items. They identified 15 tasks, including perceptual speed, complex coordination, compensatory tracking, kinesthetic memory, route walking, selective attention, time sharing, encoding speed, mental rotation, item recognition, immediate/delayed memory, decision making speed, probability estimation, risk taking, and embedded figures.

The advent of personal computers (PCs) in the 1980s greatly facilitated the implementation of these performance tasks. In the 1970s, perceptual speed and reaction time were typically measured with a tachistoscope, an instrument psychologists were happy to leave behind when PCs became powerful enough for operational use. The Computerized Adaptive Testing -- Armed Services Aptitude Battery (CAT-ASVAB) was largely developed in the 1980s and finally implemented for enlistment testing in 1993. The computer platform, state of the art at that time, consisted of an IBM PC/AT compatible running an Intel 80386 microprocessor at 25 megahertz with 640 kilobytes of memory and an additional 3 megabytes of extended memory (Unpingco, Hom, & Rafacz, 1997).

In the late 1980s the Air Force began using the Basic Attributes Test (BAT), a computer-based assessment, in addition to the Air Force Officer Qualifying Test (AFOQT). The BAT included a rotary pursuit task called Two-Hand Coordination, a control precision and multilimb coordination task called Complex Coordination, a measure of reaction time and rate control called Time Sharing, assessments of information processing capacity called Mental Rotation and Item Recognition, and an evaluation of attitude toward risk called the Activities Interest Inventory (Carretta & Ree, 1993). The combination of the BAT and the AFOQT, called the Pilot Candidate Selection Method (PCSM), was operationally implemented in 1993.

The Air Force developed the Test of Basic Aviation Skills (TBAS) in the early 2000s as an updated and enhanced replacement for the BAT (Carretta, 2005). Carretta described the TBAS assessments:

- Three Digit Listening Test, which presents a series of numbers and letters via headphones. Examinees respond when any of three identified numbers (i.e., the "targets") is presented and they are not to respond to any other number or letter (i.e., to "non-targets").
- Five Digit Listening Test, which is the same as the Three Digit Listening Test, except that there are five targets.
- Airplane Tracking Test, which is a compensatory tracking task that assesses the ability to track a target moving in two dimensions.
- Horizontal Tracking Test, which measures an examinee's ability to track a target on the horizontal axis.
- Airplane Tracking and Horizontal Tracking Test, which requires examinees to perform the Airplane Tracking Test and the Horizontal Tracking Test simultaneously.
- Airplane Tracking, Horizontal Tracking, and Three Digit Listening Test, which requires examinees to perform these three tasks simultaneously.
- Airplane Tracking, Horizontal Tracking, and Five Digit Listening Test, which requires examinees to perform these three tasks simultaneously.
- Emergency Scenario Test, which requires examinees to perform the Airplane Tracking Test and the Horizontal Tracking Test simultaneously, while also responding to an emergency situation indicated by an audio signal.
- Unmanned Aerial Vehicle (UAV) Test, in which an airplane is shown flying in a given direction with a map of the ground view. Examinees must identify map locations. For example, the computer monitor might indicate that the airplane is flying toward the Northeast and the examinee is asked to identify the south parking lot for a building.

The Navy has long been interested in pilot selection. The ASTB was originally introduced in 1942, with revisions in 1953, 1971, 1992, and 2004 (Williams, Albert, & Blower, 1999) and the introduction of parallel forms of the 1992 test in 2004. The 1992 form of the ASTB was a paper-and-pencil assessment with six subtests: math-verbal, mechanical comprehension, spatial apperception, aviation and nautical information, biographical inventory, and aviation interest. These tests are combined to form the Academic Qualification Rating, which is used to predict academic performance in ground school, the Pilot Aptitude Rating, which predicts flight grades in primary flight training, and the Pilot Biographical Inventory, which is designed to predict attrition during primary flight training.

Williams et al. (1999) also described the Computer-based Performance Test (CBPT), which includes ten assessments of tracking and information processing that are given in single- and dual-task contexts. Williams et al. described the tests:

- Two-dimensional Tracking (2-DT) task, which requires examinees to use a joystick to keep a cursor centered on crosshairs shown in the middle of the computer monitor. The cursor is "continuously driven by horizontal and vertical disturbance functions that work to displace the cursor from the center" (p. 18-3).
- Dichotic Listening (DL) task, which requires examinees to listen selectively to the information presented in one ear while additional information is presented simultaneously in the other ear.
- 2-DT and DL, which requires examinees to perform both tasks simultaneously.
- Horizontal Tracking (HT) task and 2-DT, which requires examinees to keep a cursor centered using rudder pedals while simultaneously performing the 2-DT task.

- DL, HT, and 2-DT, which requires examinees to perform all three simultaneously.
- Vertical Tracking (VT), 2-DT, and HT, where another cursor moves only vertically along the left side of the computer monitor and the examinee must keep it centered using a second joystick on the left side of the test station. The examinee must simultaneously perform the 2-DT and HT tasks.
- One-dimensional Tracking (1-DT), where the examinee is asked to keep a cursor centered on a horizontal line using the right joystick to move the cursor to the right and the left joystick to move the cursor to the left. The cursor is continuously moved off center by a disturbance function.
- Working Memory (WM), which requires examinees to "calculate the absolute difference between single digit numbers that are sequentially presented on the computer monitor" (p. 18-3).
- HT and WM, which requires examinees to perform these two tasks simultaneously.
- Manikin Test (MT), in which drawings of a sailor holding a red square in one hand and a green circle in the other appear on the computer monitor. Examinees are required to determine which of the sailor's hands is holding the red square.

The Navy's PBM Test suite consists of seven timed, interactive, performance-based subtests:

- **Directional Orientation Test (DOT).** This subtest consists of 48 discrete four-option trials. Each trial consists of two images, with the left image being an aerial view of a map depicting an aircraft oriented on a specific heading, and the right image being a forward-facing view from that aircraft, depicting a building surrounded by four parking lots oriented at right angles to each other. These two stimuli are accompanied by audio instructions to select a specific parking lot (e.g., “north”). The examinee’s task is to correctly identify the target parking lot and select it by using the joystick to position the cursor over it and depressing the trigger. The test records examinee response option selected as well as reaction time.

- **Dichotic Listening Test (DLT)**. This subtest consists of four 120 second trials, during each of which two distinct strings of letters and numbers are presented simultaneously to each ear, with a different letter or number presented to each ear at the same moment throughout the trial. Before each trial, the examinee is audibly prompted to attend to either the right or left ear. The task is to depress the trigger on the joystick grasped by the right hand when an even number is presented to the target ear, and to push a thumb trigger on the throttle grasped with the left hand when an odd number is presented to the target ear. The examinee is expected to ignore all stimuli presented to the non-target ear, as well as letters presented to the target ear. The target ear changes with each trial. During each trial, a total of 4 stimulus numbers requiring a response are presented to the target ear, embedded in a string of 8 letters.

- **Vertical Tracking Test (VTT)**. This subtest includes a graphical depiction of an aircraft approximately 0.5 inches in height and width when displayed on a 15” computer monitor at 800x600 pixel resolution, along with yellow crosshairs of approximately the same size. The examinee’s task is to place and keep the crosshairs atop the airplane throughout the trial by moving the throttle with his or her left hand. The aircraft moves up and down

randomly, with relatively constant speeds across three time intervals of 20 seconds each. During the first time interval, the speed of aircraft movement is 40 pixels per 35ms (SLOW SPEED), but it increases to 75 pixels per 35ms for the second time interval (MEDIUM SPEED) and increases again to 90 pixels per 35ms for the third interval (FAST SPEED). On this and all other tracking tasks discussed, the Euclidean distance between the cursor and the target is captured every 400ms. Additionally, during this tracking task, cursor position is compared to the target position every 35ms. When these positions coincide, a counter is incremented. When this count reaches 40, a redirect occurs, whereby the aircraft's direction changes randomly. After each redirect, the count is reset to zero. Euclidian distances and the number of redirects for each time interval are recorded in the database. In addition, the "Average Distance" score is computed as the mean of all 400ms intervals captured during the test. The VTT is preceded by a 30s practice session.

- **Airplane Tracking Test (ATT)**. The task is very similar to the VTT, but the aircraft moves on two axes, permitting movement in any direction on the plane of the monitor. The aircraft moves with variable speeds across three time intervals of 20 seconds each. During the first time interval, the speed of the aircraft ranges between 115 and 203 pixels per 35ms (SLOW SPEED); during the second time interval, the speed of the aircraft ranges between 150 and 265 pixels per 35ms (MEDIUM SPEED); and during the third time interval, the speed of the aircraft ranges between 175 and 309 pixels per 35ms (FAST SPEED). The speeds within each time interval are randomly generated each time the target changes direction (e.g., with each redirect). The crosshair is controlled by moving the joystick with the examinee's right hand. The only other difference between this task and the VTT is that on this task, a redirect takes place when the coincident position counter reaches 30. As with the VTT, all Euclidian distances between the crosshair and the target are recorded for each of the three aircraft speeds. The number of redirects and the average distance score are also recorded. The ATT is preceded by a 30s practice session.

- **Airplane/Vertical Tracking Test (ATTVTT)**. This subtest requires the examinee to perform the ATT and VTT simultaneously across three time intervals of 40 seconds each. The speed increments and intervals are identical to those for the VTT and ATT subtests presented alone, as are the algorithms for updating the target position in the VTT and ATT components. Each examinee tracks the vertically moving aircraft target with the throttle on the left side of the computer and the freely moving target with the joystick on the right side of the monitor. The ATTVTT is preceded by a 30s practice session.

- **Multi-Tracking Test (MTT)**. This subtest requires the examinee to perform the ATT and VTT tasks together with the dichotic listening task (DLT) across three time intervals of 60 seconds each. The speed increments and target position update algorithms are identical to those used for the VTT and ATT subtests presented alone. The MTT is preceded by a 30s practice session.

- **Emergency Scenario Test (EST)**. This subtest begins with the presentation of two different three-step emergency procedures involving examinee-entered changes to fuel

flow, engine power, and propeller position using buttons on the joystick and throttle. Examinees are presented written instructions identifying the conditions under which these procedures are to be followed, after which the examinee is asked to perform the ATT and VTT subtests simultaneously with the additional attentional demand of remembering when and how to respond to emergencies across three time intervals of 40s each. The speed increments are identical to the VTT and ATT subtests. There is no practice session for the EST.

All subtests must be completed within the allocated time to receive valid results. The examinee is required to use a joystick and throttle to complete all portions of the test.

Levels of Analysis for PBM Data

Examinee behavior on each PBM subtest provides a rich source of psychometric information:

- "item-level" information that is used to compute subtest-level indicators (see Table 1)
- 35 subtest-level variables that are recorded in the PBM results database during every testing session
- composites of subtest-level variables

For example, the DOT records the accuracy and latency of examinee responses for each of 48 direction orientation trials and that information is used to compute variables such as the total correct, total incorrect, and the cumulative response time associated with those answers. Because the DOT trials are independent of each other and the same across examinees, the data for this subtest can be examined in a relatively straightforward manner both at the subtest level and at an item (trial) level using classical test theory (CTT) and/or item response theory (IRT) methods.

The use of low-level information captured during other subtests poses greater challenges for psychometric analyses because stochastic design features that introduce "controlled variation" across examinees inherently change the particulars of the task at any point in time and introduce noise into the data. In such cases, the items are essentially slices of time (e.g., readings taken every 400ms) and the item responses are, for example, the Euclidian pixel distances between the examinee-controlled crosshairs and an airplane moving in one or two dimensions. Higher level data include the number of controlled redirects that occur during an exercise because an examinee has the crosshairs centered on a target when a reading is taken. Whether or not the low-level data provided information beyond the higher-level descriptors was a central question in this investigation.

Another interesting feature of the PBM that has implications for scoring is the retention of common elements across stages of the assessment. The seven subtests of the PBM were created and arranged in a sequence designed to increase the cognitive load and, accordingly, the stress on the examinee as he/she progresses toward the final subtest (EST) involving emergency scenarios. EST retains elements of several previous subtests, requiring an examinee to respond to three emergencies of increasing difficulty, by manipulating controls on a throttle and a joystick and simultaneously performing an ATTVTT exercise, which in turn has elements in common with VTT and ATT. Thus, rather than focusing solely on the successful resolution of the three emergency scenarios when scoring EST, one can also examine performance data for the different elements and take those into account when developing subtest scores. Moreover, one can form high level composites by adding performance scores for dichotic listening, one-dimensional tracking, and two-dimensional tracking tasks *across the seven subtests* to see whether such high-level composites can be combined with scores on the DOT to predict criterion performance as well or better than the seven individual subtest scores.

This report therefore presents results for three "levels of analysis" wherever possible. Subtest-level scores were created using information readily available from the 35 indicators shown in Table 1. In addition, where IRT models could be applied to lower-level data, latent trait estimates (examinee scores) derived from those analyses were examined to determine if better criterion-related validities resulted when using them in lieu of other subtest-level indicators. Finally, high-level composites were formed by adding scores for the common elements across subtests. In

addition, the validity of these high-level predictors were compared to the individual subtest scores derived by classical test theory and, where applicable, IRT methods. The results of these analyses follow a brief description of the subtest-level indicators, examinee demographics, and criterion information presented next.

Table 1. Readily Available PBM Subtest-Level Variables and Their Descriptions

Count	PBM Subtest	Variable Name	Description
1	Directional Orientation Test (DOT)	DOTTotalQuestions	Total number of Direction Orientation questions presented to examinee.
2		DOTTotalCorrect	Number of Direction Orientation questions the examinee answered correctly.
3		DOTTotalCorrectTime	The sum of the examinee's response times to all questions that he or she answered correctly.
4		DOTTotalIncorrect	Number of items that the examinee answered incorrectly.
5		DOTTotalIncorrectTime	The sum of the examinee's response times to all questions that he or she answered incorrectly.
6		DOTTotalTime	The sum of the examinee's response times to all questions.
7	Dichotic Listening Test (DLT)	DLTTotalQuestions	Number of target trials in the dichotic listening task.
8		DLTTotalCorrect	Number of target trials on which the examinee responded correctly.
9		DLTTotalCorrectTime	Sum of response times on DLT trials in which the examinee responded correctly.
10		DLTTotalIncorrect	Number of target trials on which the examinee responded incorrectly.
11		DLTTotalIncorrectTime	Sum of response times to DLT target trials in which the examinee responded incorrectly.
12		DLTTotalTime	Sum of response times to target trials on DLT regardless of accuracy.
13	Airplane Tracking Test (ATT)	ATTAvgDistance	Mean distance between target and cursor over all 400ms intervals of entire airplane tracking task.

14		ATTRedirects	Number of times that the target was redirected during the airplane tracking task because the examinee had the cursor over the target.
15	Vertical Tracking Test (VTT)	VTTRedirects	Number of times that the target was redirected during the vertical tracking task because the examinee had the cursor over the target.
16		VTTAvgDistance	Mean distance between target and cursor over all 400ms intervals of entire vertical tracking task.
17	Airplane/Vertical Tracking Test (ATTVTT)	AttVtt_VTTRedirects	Number of times that the target of the vertical tracking task was redirected because the examinee had the cursor over the target during the combined airplane tracking and vertical tracking task.
18		AttVtt_VTTAvgDistance	Mean distance between target of the vertical tracking task and cursor across all 400ms intervals of the combined airplane tracking and vertical tracking task.
19		AttVtt_ATTRedirects	Number of times that the target of the airplane tracking task was redirected because the examinee had the cursor over the target during the combined airplane tracking, and vertical tracking task.
20		AttVtt_ATTAvgDistance	Mean distance between target of the airplane tracking task and cursor across all 400ms intervals of the combined airplane tracking and vertical tracking task.
21	Multi-Tracking Test (MTT)	AttVttDlt_VTTRedirects	Number of times that the target of the vertical tracking task was redirected because the examinee had the cursor over the target during the combined airplane tracking, vertical tracking, and dichotic listening task.
22		AttVttDlt_VTTAvgDistance	Mean distance between target of the vertical tracking task and cursor across all 400ms intervals of the combined airplane tracking,

		vertical tracking, and dichotic listening task.
23	AttVttDlt_ATTRedirects	Number of times that the target of the airplane tracking task was redirected because the examinee had the cursor over the target during the combined airplane tracking, vertical tracking, and dichotic listening task
24	AttVttDlt_ATTAvgDistance	Mean distance between target of the airplane tracking task and cursor across all 400ms intervals of the combined airplane tracking, vertical tracking and dichotic listening tasks.
25	AttVttDlt_DLTTotalQuestions	Number of target trials in the dichotic listening task during the combined airplane tracking, vertical tracking, and dichotic listening task.
26	AttVttDlt_DLTTotalCorrect	Number of target trials on which the examinee responded correctly during the combined airplane tracking, vertical tracking, and dichotic listening task.
27	AttVttDlt_DLTTotalCorrectTime	Sum of response times on DLT trials in which the examinee responded correctly during the combined airplane tracking, vertical tracking, and dichotic listening task.
28	AttVttDlt_DLTTotalIncorrect	Number of DLT target trials on which the examinee responded incorrectly during the combined airplane tracking, vertical tracking, and dichotic listening task.
29	AttVttDlt_DLTTotalIncorrectTime	Sum of response times to DLT target trials in which the examinee responded incorrectly during the combined airplane tracking, vertical tracking, and dichotic listening task.
30	AttVttDlt_DLTTotalTime	Sum of response times to target trials on DLT during the combined airplane tracking, vertical tracking, and dichotic listening task,

			regardless of accuracy.
31	Emergency Scenario Test (EST)	AttVttScn_VTTRedirects	Number of times that the target of the airplane tracking task was redirected because the examinee had the cursor over the target during the emergency scenario.
32		AttVttScn_VTTAvgDistance	Mean distance between target and cursor on the vertical tracking task across all 400ms intervals of the emergency scenario.
33		AttVttScn_ATTRedirects	Number of times that the target on the airplane tracking task was redirected because the examinee had the cursor over the target during the emergency scenario.
34		AttVttScn_ATTAvgDistance	Mean distance between target and cursor on the airplane tracking task across all 400ms intervals of the emergency scenario task.
35		AttVttScn_EndingSkill	Number of emergencies the examinee correctly responded to during the emergency scenario task.

DEMOGRAPHIC AND CRITERION DATA

The data set used in this report contained complete PBM data for 310 Student Pilots and 89 Student Naval Flight Officers. In addition to PBM, a variety of demographic and criterion data were available.

Demographics

The sample consisted of 310 Student Pilots (SPs) and 89 Student Naval Flight Officers (SNFOs). The majority of examinees were college graduates (91.8 %), male (94.0 %), and Caucasian (90.2). There were 20 Hispanics, 6 African Americans, 12 Asians, and 1 Native American in the sample. Because the number of minority examinees was low, no gender- or race-based analyses were performed.

More than half of the sample was composed of US Navy officers (59%), followed by US Marine Corps officers (36%) and US Coast Guard officers (5%). Frequencies indicating the present military status of examinees are shown in Table 2.

Table 2. Present Military Status Statistics

Present Status	Frequency	Percent
Officer, US Coast Guard	19	4.8
Officer, US Marine Corps	144	36.1
Officer, US Navy	236	59.1
Total	399	100.0

Prior Flight Experience, Simulator Experience and ASTB Scores

In addition to demographic information, data for several variables relevant to students' PBM test performance were also available. These included ASTB subtest and composite scores, prior simulator experience, and number of hours of prior flight experience. Because ASTB composition and scoring was changed in 2004, students who took ASTB prior to that date were excluded from the analyses (N = 67). Various ASTB scores and composites can be used as statistical control variables for evaluating PBM's incremental contribution to the prediction of training performance. The other two variables, simulator experience and hours of prior flight experience were used as indicators of prior relevant training and therefore expected to be positively related to PBM test scores. These variables were used to investigate the construct validity of PBM scores. Note that because the hours of prior flight experience variable was severely skewed, its values were recoded into 6 categories where l = zero hours, 2 = .10 to 15 hours … and 6 = 300 to 10000 hours (See Table 5, below, for the complete list).

Table 3 shows the correlation matrix of the ASTB subtests and composites based on the sample of 332 students who had taken the battery after 2004. Because this sample had been pre-selected based on the ASTB scores, correlations between ASTB subtests were not very high. Nevertheless, Mathematical, Mechanical, and Reading comprehension subtests correlated .24 to .41, which indicates the presence of a general cognitive ability factor. As expected, the Aviation and Nautical Information and Spatial Apperception subtests had lower correlations with other subtests. The four ASTB composites correlated .62 to .89 with each other.

Table 3. Correlations Between ASTB Subtests and Composites

	ANI	MST	RCT	SAT	MCT	AQR	PFAR	FOFAR	OAR
Subtest: Aviation and Nautical Information (ANI)									
Subtest: Mathematical Comprehension (MST)	-.04								
Subtest: Reading Comprehension (RCT)	.11	.31							
Subtest: Spatial Apperception (SAT)	.18	.08	.15						
Subtest: Mechanical Comprehension (MCT)	.19	.41	.24	.33					
Academic Qualification Rating Composite (AQR)	.56	.61	.40	.45	.82				
Pilot Flight Aptitude Rating Composite (PFAR)	.78	.27	.26	.61	.66	.89			
Flight Officer Aptitude Rating Composite (FOFAR)	.44	.69	.46	.66	.53	.87	.79		
Officer Aptitude Rating Composite (OAR)	.14	.72	.39	.29	.93	.88	.62	.70	

Note: Correlations higher than .10 are significant ($p < .05$).

Statistics for the prior simulator experience and hours of prior flight experience are presented in Tables 4 and 5, respectively. As can be seen in Table 4, the range of hours of prior flight experience was quite large in the examinee sample. About 81 % of the examinees reported either no simulator experience or just enough to be declared as novices, whereas 19 % were classified as intermediate or expert. A similar pattern was observed in the prior flight hours data, where 70.4 had no prior flight experience as shown in Table 5.

Table 4. Prior Simulator Experience Frequency Statistics

Simulator Experience	Frequency	Percent
None	190	47.62
Novice	133	33.33
Inter	64	16.04
Expert	12	3.01
Total	399	100.00

Table 5. Hours of Prior Flight Experience Frequency Statistics

Flight Hours	Frequency	Percent
None	281	70.4
.1 to 15	35	8.8
16 to 29	46	11.5
20 to 59	8	2.0
60 to 99	11	2.8
100 to 299	11	2.8
300 to 10000	7	1.8
Total	399	100.00

Criterion Data

The criterion data against which the PBM test scores were validated consisted of students' scores in Primary phase flight training (ground training scores were excluded here). The curriculum for Primary phase flight training consists of four stages of interest: Contact, Instrument, Navigation, and Formation training. Each stage consists of multiple blocks that pertain to different content or instructional goals, and within each block are a series of events for which students receive grades. All blocks are identified by a three-digit code consisting of a letter identifying the block's stage followed by a two-digit code. If the first numeral of this two-digit code is a 2, all events of the block are performed in a flight simulator. If this number is a 4, all block events are performed in an aircraft.

Contact Stage:

The purpose of the contact stage is to familiarize the student with the aircraft, its systems and their operation, common emergencies, and fundamental aviation procedures under visual flight rules.

Instrument Stage:

The instrument stage focuses on operation of the aircraft and navigation under instrument flight rules, wherein the student is required to be able to safely operate and navigate the aircraft without reliance on visual cues from outside the cockpit.

Navigation Stage:

In this stage, students are required to successfully plan, navigate, and execute a low-altitude (between 1,000 - 3,000 feet above ground level [AGL] for daytime flight and between 2,000 – 4,500 feet AGL at night) overland flight to a different airfield with a specific arrival time using only a chart, visual references, speed, heading, and time. Students are prohibited from using navigational aids.

Formation Stage:

This stage introduces the student to flight operations in a two-aircraft section. Students practice both cruise (larger separation between aircraft) and parade (closer interval) formation flight. In this sample, very few student pilots completed the solo flight in the F41 block, so no analyses were performed for this particular training segment. Finally, block F43 is flown only by US Air Force students participating in Navy Flight Training, so none of the students in this study had data on this criterion.

SPs and SNFOs must meet different training requirements in preparation for their respective job functions. Although they participate in similarly titled training blocks, the actual content emphasis and grading criteria differ for the two groups. For this reason, the Primary phase requirements and scores must be considered separately for SPs and SNFOs. Curriculum differences for these two groups appear in Tables 6 and 7 below.

Table 6. Primary Pilot Flight Training Curriculum Blocks for SPs

Training Block Name	Description	Training Media	Number of Events/Hours
C20	Cockpit procedure training	CPT	5 / 6.5
C40	Day contact: basics, grades not used in ranks	T-34	4 / 6.4
C41	Day contact: graded familiarization flights	T-34	4 / 7.6
C42	Day contact: graded flights with briefs	T-34	4 / 8.0
C43	Day contact check ride	T-34	1 / 2.0
C44	Initial contact solo: four touch-and-go landings	T-34	1 solo / 1.5
C45	Day contact: aerobatics	T-34	3 + 2 solo / 9.0
C46	Day contact: aerobatics 2	T-34	1 + 1 solo / 3.6

C47	Night contact	T-34	2 / 3.0
I20	Basic instruments: introduction	IFT	4 / 5.2
I21	Basic instruments: emergencies	IFT	3 / 3.9
I40	Basic instruments: spatial disorientation demonstration	T-34	3 / 4.5
I22	Radio instruments: introduction to radar equipment	IFT	5 / 6.5
I23	Radio instruments: real world and emergency situations	IFT	4 / 5.2
I41	Radio instruments: graded flights	T-34	5 / 9.0
I24	Instrument navigation: real time locals	IFT	6 / 7.8
I25	Instrument navigation: real time out-and-ins	IFT	4 / 5.2
I42	Instrument navigation: 1+ high altitude and 1+ night flight	T-34	4 / 8.0
I43	Instrument stage check ride	T-34	1 / 2.0
N40	Day navigation	T-34	2 / 3.2
N41	Night navigation	T-34	2 / 3.2
F40	Basic formation	T-34	5 / 10.5
F41	Basic formation solo	T-34	1 solo / 1.5
F42	Cruise formation	T-34	3 / 6.0
F43	Air Force formation	T-34	6 / 12.0

Notes: CPT = Cockpit Procedures Trainer, a flight simulator with no moving parts or powered gauges, IFT = Instrument Flight Trainer, a flight simulator with powered gauges, but no visual depiction of the environment outside the cockpit. The T-34 is a fixed-wing propeller-driven aircraft.

Table 7. Primary Pilot Flight Training Curriculum Blocks for SNFOs

Training Block Name	Description	Training Media	Number of Events/Hours
C20	Cockpit procedure training	UTD/OFT	3 / 4.5
C40	Day contact: preflight briefings and basic procedures	T-6A	4 / 6.0
C41	Night contact	T-6A	1 / 1.5
C42	Day contact check ride	T-6A	1 / 1.5
I20	Instrument navigation: introduction	UTD/OFT	9 / 13.5
I40	Instrument navigation: basic operations	T-6A	5 / 10.0
I41	Instrument navigation check ride 1	T-6A	1 / 2.0

I42	Instrument navigation: emergency procedures 1	T-6A	4 / 8.0
I43	Instrument navigation: emergency procedures 2	T-6A	4 / 8.0
I44	Instrument navigation check ride 2	T-6A	1 / 2.0
N30	Day visual navigation: introduction	OFT	2 / 3.0
N50	Day visual navigation: VFR between 1000 and 3000 feet AGL	T-6A	5 / 10.0
N51	Visual navigation check ride	T-6A	1 / 2.0
F50	Formation: responsibilities, positions, and procedures	T-6A	2 / 3.5
F51	Formation navigation: two-ship navigation procedures	T-6A	2 / 4.0

Notes: UTD = Undergraduate Training Device, a flight simulator with no moving parts or powered gauges, OFT = Operational Flight Trainer, a flight simulator with powered gauges and a visual depiction of the environment outside the cockpit. The T-6A is a fixed-wing ejection-seat propeller-driven aircraft.

Reporting Training Grades:

On each simulator or flight event, a student pilot is awarded between 10 and 30 grades on specific maneuvers or tasks using a four-point Likert scale for each grade. These grades are compared to a minimum standard on the same scale defined for each maneuver in the curriculum. Grades awarded are divided by the required performance standard for each maneuver or task attempted during a training event, block, or set of blocks to yield a raw score for that interval of training.

Students are also awarded an overall categorical grade for each simulator and flight event; available options are *pass* (coded as 0), *marginal* (coded as 0.5), or *unsatisfactory* (coded as 1.0). The sum of the overall grades that are awarded for a training interval accounts for 10% of the student's point total for that interval. The raw scores for the tasks and maneuvers described in the paragraph above accounts for the other 90% of this total. These two grades are then normed and summed. However, because the overall event grades typically exhibit a strong negative skew, this sum is then normed again and scaled as a T-score, with mean = 50.0 and SD = 10.0.

This T-score, referred to as a student's Navy Standard Score (NSS), is calculated for each block within the curriculum, as well as for the set of blocks that student has completed to date. This latter NSS value is referred to as a student's interim NSS in cases where the student has completed only a portion of the curriculum, and his or her overall NSS in cases where he or she has completed the entire curriculum. The norm group used to define a block or interim NSS only includes data from the most recent 60 students to complete that block or set of blocks, ignoring any data from these students on blocks beyond the set included in a specific interim norm group. Thus, it is possible in a 25 block curriculum such as the Navy Primary Pilot Flight Training, to require definition of norms on 25! (1.55×10^{25}) distinct norm groups, although the number of active groups at any given time is usually less than 50 due to the patterns by which students typically progress through blocks of training.

Note that on solo flights, students are typically awarded no task or maneuver grades, save under exceptional circumstances such as, the occurrence of a mishap or an overt safety violation. It is therefore unusual for students to receive NSSs for blocks C44 and F41, which consist of one solo flight each. Blocks C46 and C47 typically include one or more solo flights each, making NSS values for these blocks less common as well.

Table 8 presents descriptive statistics for the T-scaled NSS values for each of the 24 blocks in the Primary Pilot Flight Training curriculum, as well as the overall NSS for the entire curriculum. As can be seen in the table, criterion data for SPs were available for 23 training blocks. For the SNFOs, data were available for the 9 out of 24 training blocks.

Table 8. Descriptive Statistics for the 24 Individual Training Criterion T-Scores and the Overall NSS

Training Block Name	Student Pilots			Student Naval Flight Officers		
	N	Mean	SD	N	Mean	SD
C20	310	49.4	9.8	89	48.3	8.3
C40	-	-	-	86	48.8	11.1
C41	292	49.2	9.4	82	51.1	9.8
C42	284	48.9	10.1	83	50.3	10.4
C43	270	49.2	9.6	-	-	-
C44	**4**	**46.4**	**18.2**	-	-	-
C45	262	50.2	9.2	-	-	-
C46	246	50.3	9.6	-	-	-
C47	238	50.7	8.9	-	-	-
I20	303	49.1	10.5	84	49.6	9.4
I21	300	49.4	10.1	-	-	-
I22	207	50.0	9.5	-	-	-
I23	206	49.7	9.9	-	-	-
I24	194	51.0	9.6	-	-	-
I25	188	51.3	9.4	-	-	-
I40	298	49.3	10.8	82	50.0	9.9
I41	200	50.6	9.3	82	49.8	10.0
I42	183	50.7	9.6	50	50.6	10.4
I43	178	50.9	9.3	49	48.9	9.8
F40	225	50.6	9.5	-	-	-
F41	**15**	**53.2**	**12.0**	-	-	-
F42	209	50.3	10.2	-	-	-
N40	183	50.8	9.1	-	-	-
N41	181	50.2	10.0	-	-	-
Navy Standard Score (NSS)	310	49.2	9.7	89	48.5	10.9

Note: Bold values represent blocks with solo flights where grades are not typically given.

Because the C44 and F41 blocks had very small sample sizes, they were dropped from subsequent analyses. Also note that because grades from individual blocks of the curriculum represented relatively short intervals of student performance and were therefore more likely to be unreliable, grades for the 22 retained blocks were aggregated into criterion composites corresponding to their respective curriculum stages (4 stages). Contact and Instrument criterion composites were also split by training medium (i.e., simulation vs. aircraft) to form more refined performance indicators. Finally, the Instrument stage was also split into Basic, Radio, and Navigation composites.

To form each composite, grades from relevant blocks were weighted by the number of events a specific student has participated in, summed, and then divided by the number of total events for that student. Hence, grades from training blocks with more events were more influential than those with a smaller number of events.

Table 9 shows the resulting means, standard deviations, reliabilities, and intercorrelations for the eleven Navy Pilot Primary Flight Training criterion composites as well as the overall curriculum NSS grade.

Table 9. Descriptive Statistics and Correlations for Navy Pilot Flight Training Weighted Criterion Composites and the Overall Navy Standard Score

	Criterion Composite	N	Min.	Max.	Mean	SD	1	2	3	4	5	6	7	8	9	10	11	12
1	Contact_Simulation	399	20.0	72.6	49.1	9.5		.42	.72	.52	.40	.53	.50	.47	.49	.17	.34	.64
2	Contact_AIRCRAFT	378	20.0	73.8	49.1	7.9	.42		.93	.68	.61	.73	.69	.63	.68	.32	.58	.82
3	Contact_ALL	399	20.4	68.6	49.2	7.3	.72	.92		.70	.61	.73	.70	.67	.72	.32	.59	.88
4	Instruments_Simulation	387	20.8	73.9	49.2	9.0	.49	.64	.67		.60	.96	.88	.88	.89	.28	.58	.86
5	Instruments_AIRCRAFT	380	20.0	75.6	48.9	8.9	.41	.56	.58	.57		.79	.74	.81	.80	.37	.51	.71
6	Instruments_ALL	387	20.8	69.1	49.1	8.0	.52	.67	.70	.93	.81		.92	.90	.91	.32	.61	.88
7	Instruments_BASIC	387	20.8	74.9	49.2	8.7	.49	.65	.67	.88	.74	.93		.69	.70	.22	.57	.83
8	Instruments_RADIO	289	20.0	80.0	50.0	8.9	.42	.52	.56	.67	.72	.77	.60		.71	.24	.55	.83
9	Instruments_NAVIGATION	244	21.9	79.1	50.6	8.3	.46	.61	.66	.76	.82	.88	.66	.64		.38	.55	.85
10	Navigation_AIRCRAFT	183	24.0	80.0	50.4	8.6	.17	.32	.32	.28	.37	.32	.22	.24	.38		.20	.34
11	Formation_AIRCRAFT	225	23.6	71.8	50.2	8.9	.34	.58	.59	.58	.51	.61	.57	.55	.55	.20		.69
12	Navy Standard Score (NSS)	399	20.0	80.0	49.1	10.0	.63	.79	.86	.81	.74	.88	.82	.73	.84	.34	.69	

Note: Correlations below diagonal are for the full sample. Correlations above diagonal are for Students Pilots only.

DIRECTIONAL ORIENTATION TEST (DOT): SCORING STRATEGIES AND VALIDITIES

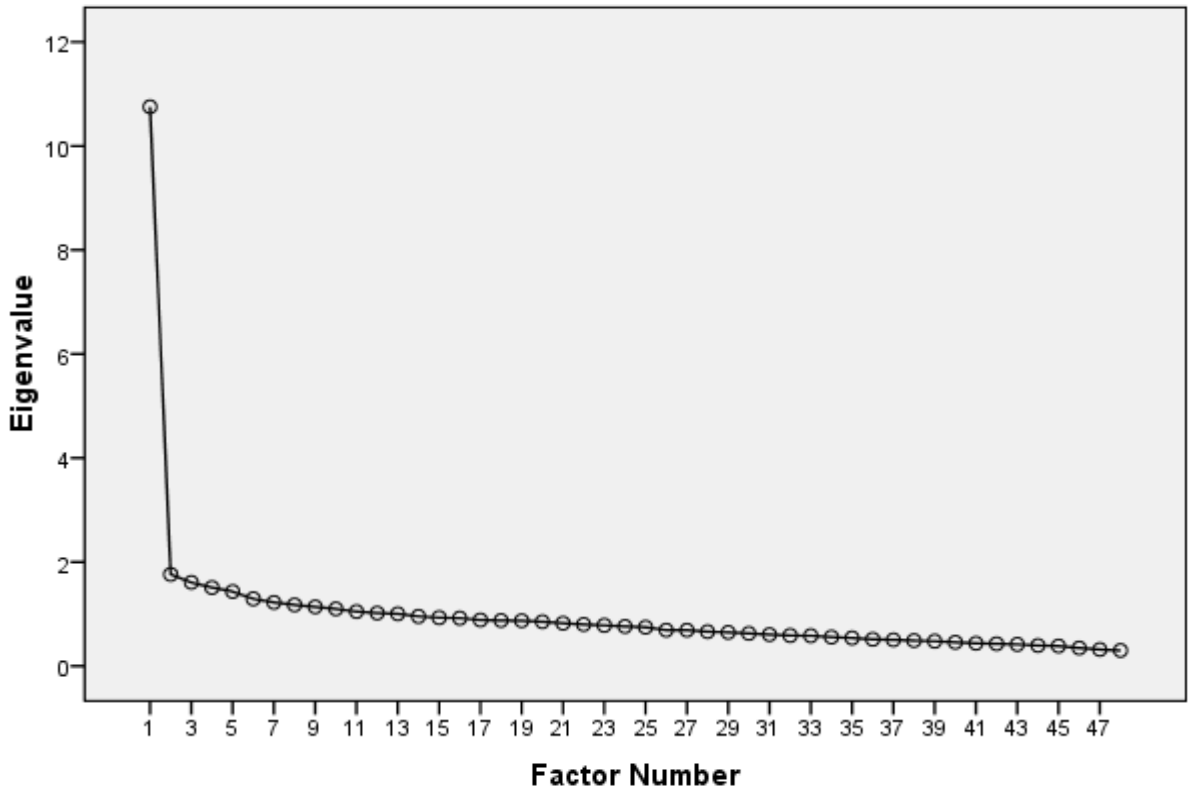
The DOT consists of 48 discrete trials involving four possible examinee responses, only one of which is correct. Each trial requires an examinee to rapidly process two visual stimuli: a map depicting an aircraft on a specific heading and a forward-facing view from that aircraft showing a building surrounded by four parking lots situated at right angles to each other. The examinee must respond to aural instructions to "image" a designated parking lot (e.g., north) by using a mouse. If the examinee correctly identifies the target parking lot, he/she receives a score of "1" for that trial; otherwise he/she is assigned a score of "0". Examinee response time is recorded for each trial as well.

DOT data were analyzed in several steps. First, we investigated the psychometric properties of individual items using principal component and classical test theory (CTT) methods. Next, item response theory (IRT) analyses were conducted. Because DOT items have 4-possible response options and only one of these options is correct, the three-parameter logistic (3PL) IRT model was fit to the DOT data. These analyses provide an important foundation for future differential item and differential test functioning analyses. Finally, we examined the criterion-related validities of the DOT by correlating its scale scores and response times with individual training grades and training composites. Although IRT and total test scores correlate highly, results for both sets of scores are reported to see which yield higher validities. We also conducted several regression analyses to see how different DOT score components predict training criteria for the total sample as well as for student pilots only.

Item-Level CTT and IRT Analyses and Results for the DOT

IRT analyses using commonly available models require that the response data can be accounted for by a single dominant dimension. Because the response data were scored dichotomously and the DOT essentially measures cognitive abilities, we chose the three-parameter logistic model (3PLM; Birnbaum, 1968) as the basis for item analysis. To check the unidimensionality assumption of the 3PLM, we ran a principal component (PCA) analysis on the inter-item correlations and plotted the eigenvalues to produce the scree plot in Figure 1. As can be seen in the figure, the ratio of first to second eigenvalues well exceeded 3.0 and there was a smooth tail formed by the second and subsequent eigenvalues, signifying that the response data were sufficiently unidimensional for the application of the 3PLM (Drasgow & Parsons, 1983; Lord, 1980).

Figure 1. Scree Plot for the Principal Component Analysis of the 48 DOT Items



IRT Calibration of the 48 DOT Items

The 3PL model was fit to the DOT data. Here, the probability of a correct response to the i^{th} item, $P_i(\theta)$, is given by

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-Da_i(\theta - b_i)]},$$

where a_i is the item discrimination parameter, b_i is the item difficulty parameter, and c_i is the lower asymptote (i.e., pseudo-guessing) parameter for item i, and D is a constant set equal to 1.702 so that the scaling of the 3PL model closely matches that of the normal ogive model.

The BILOG-MG computer program (Zimowski, Muraki, Mislevy, & Bock, 1996) was used to estimate 3PLM item parameters. The input file used to estimated 3PLM item parameters is shown below.

```
>3pl parameters DOT399
>PBM DOT data data
>GLOBAL DFName = 'DOT399.dat',
```

```
      NPArm = 3, SAVE;
>SAVE    PARM='DOT399.PAR',
       COV = 'DOT399.COV',
       SCO = 'DOT399.sco',
       ISTAT = 'DOT399.ctt';
>LENGTH NITems = (48);
>INPUT   SAMPLE=99999, NIDCHAR=6,
       NFNAME='notrch.key';
>ITEMS ;
>TEST1 TNAme = 'DOT399',
     INUmber = (1(1)48);
(6A1, t1, 48A1)
>CALIB   NQPT=40, CYCLES=100, NEWTON=35, CRIT=0.01, PLOT=0; NOFLOAT;>SCORE ;
```

Model-data fit was evaluated using both graphical methods (fit plots) and statistical methods (adjusted chi-square to degrees of freedom ratios for individual items (singlets), pairs of items (doublets), and groups of three items (triplets), as suggested by Drasgow, Levine, Tsien, Williams, & Mead (1995). These analyses were performed using the MODFIT-Z 2.0 computer program (Stark, 2007). Overall the fit plots indicated that the 3PLM fit the DOT response data well. This finding was confirmed by the chi-square analyses (Table 10), which yielded means of .03, 1.15, and 1.57 for singlets, doublets, and triplets respectively. In general, adjusted chi-square to degrees of freedom ratios of less than 3 indicate a good model-data fit.

Table 10. Chi-Square Model-Data Fit Statistics for Items of the DOT

FREQUENCY TABLE OF ADJUSTED (N=3000) CHISQUARE/DF RATIOS

	<1	1<2	2<3	3<4	4<5	5<7	>7	Mean	SD
Singlets	47	1	0	0	0	0	0	0.03	0.18
Doublets	42	0	1	1	1	0	3	1.15	3.93
Triplets	11	1	1	0	0	2	1	1.57	2.49

Table 11 presents CTT statistics, IRT parameter estimates, and response time information for the 48 DOT items. Shown are the item means (P-values), standard deviations (SD), corrected item-total correlations (CITC), 3PLM discrimination (a), difficulty (b), and pseudo-guessing (c) parameters, as well as the average examinee response times and the corresponding standard deviations. Note that many of the corrected item-total correlations are fairly large (>.4), as are the IRT *a* parameter estimates, indicating that the items are quite discriminating. Moreover, the wide range of p-values and IRT *b* parameter estimates suggests that the test provides good measurement across a broad range of examinee ability.

Table 11. CTT, IRT, and Response Time Statistics for the 48 DOT Items

DOT Item Name	CTT Statistics			IRT 3PLM Parameters			Response Times	
	p-value	SD	CITC	a	B	c	Mean	SD
DOT01ACC	.53	.50	.45	1.20	0.25	0.15	6.53	6.44
DOT02ACC	.80	.40	.53	1.50	-0.64	0.26	3.76	2.99
DOT03ACC	.74	.44	.51	1.09	-0.56	0.20	4.38	2.76
DOT04ACC	.66	.48	.42	0.87	-0.22	0.20	4.63	2.78
DOT05ACC	.62	.49	.42	1.06	0.04	0.22	5.84	5.09
DOT06ACC	.90	.30	.40	0.93	-1.63	0.26	2.84	1.53
DOT07ACC	.74	.44	.36	0.81	-0.47	0.28	4.37	2.66
DOT08ACC	.84	.37	.42	0.82	-1.25	0.23	4.10	2.82
DOT09ACC	.54	.50	.34	0.87	0.39	0.21	5.35	4.43
DOT10ACC	.67	.47	.34	0.80	-0.17	0.26	4.79	2.89
DOT11ACC	.95	.21	.36	1.03	-2.35	0.22	2.42	1.35
DOT12ACC	.47	.50	.43	0.92	0.36	0.11	5.77	5.07
DOT13ACC	.76	.43	.54	1.08	-0.74	0.16	4.89	3.22
DOT14ACC	.76	.43	.57	1.13	-0.74	0.13	3.70	2.11
DOT15ACC	.70	.46	.35	0.64	-0.57	0.20	5.05	3.44
DOT16ACC	.69	.46	.52	0.95	-0.49	0.12	4.20	2.87
DOT17ACC	.52	.50	.50	0.99	0.10	0.08	5.72	4.75
DOT18ACC	.72	.45	.49	1.01	-0.50	0.19	5.09	4.21
DOT19ACC	.62	.49	.38	0.82	-0.01	0.22	6.03	4.41
DOT20ACC	.49	.50	.44	0.88	0.27	0.09	5.70	4.25
DOT21ACC	.84	.37	.34	0.61	-1.55	0.21	4.09	2.88
DOT22ACC	.83	.37	.35	0.70	-1.28	0.25	4.66	2.64
DOT23ACC	.56	.50	.44	0.85	0.06	0.13	5.24	4.05
DOT24ACC	.92	.27	.40	0.94	-2.01	0.19	2.84	1.49
DOT25ACC	.83	.38	.38	0.72	-1.29	0.22	3.83	2.58
DOT26ACC	.49	.50	.43	0.86	0.28	0.10	4.93	3.69
DOT27ACC	.75	.43	.58	1.16	-0.71	0.12	3.71	2.99
DOT28ACC	.84	.37	.38	0.74	-1.38	0.22	4.42	3.25
DOT29ACC	.96	.19	.35	1.11	-2.44	0.21	2.63	1.28
DOT30ACC	.94	.24	.28	0.70	-2.59	0.21	2.69	1.78
DOT31ACC	.71	.46	.40	0.79	-0.45	0.22	4.44	3.28
DOT32ACC	.68	.47	.37	0.61	-0.52	0.17	5.01	3.21
DOT33ACC	.64	.48	.38	0.68	-0.26	0.18	5.72	4.24
DOT34ACC	.58	.49	.46	0.83	-0.07	0.11	5.21	4.53
DOT35ACC	.62	.49	.45	0.85	-0.19	0.14	4.93	3.32
DOT36ACC	.80	.40	.57	1.11	-0.96	0.13	3.20	2.21
DOT37ACC	.77	.42	.49	0.90	-0.89	0.15	4.90	3.65
DOT38ACC	.65	.48	.44	0.79	-0.34	0.15	5.14	3.33
DOT39ACC	.89	.31	.51	1.10	-1.60	0.16	2.74	1.52
DOT40ACC	.61	.49	.43	0.73	-0.18	0.13	5.31	3.77
DOT41ACC	.78	.41	.28	0.49	-1.39	0.19	4.00	2.68
DOT42ACC	.85	.35	.48	0.93	-1.39	0.17	3.38	1.90
DOT43ACC	.82	.39	.56	1.12	-1.08	0.14	2.93	1.50

DOT44ACC	.96	.20	.27	0.78	-2.88	0.19	2.42	1.25
DOT45ACC	.76	.43	.54	0.89	-0.92	0.10	3.26	1.92
DOT46ACC	.72	.45	.43	0.72	-0.67	0.17	4.66	3.41
DOT47ACC	.74	.44	.49	0.88	-0.72	0.16	4.08	2.58
DOT48ACC	.71	.45	.40	0.65	-0.72	0.16	3.97	2.59

DOT Scale Scores

The DOT items assess two kinds of abilities/skills: spatial rotation and cognitive processing speed. The total number correct scores (DOT Total Correct) and the IRT-based trait scores (DOT IRT Score) are the best indicators of examinee spatial ability. Although these two indicators are highly correlated (r = 0.98), the DOT IRT Score is a weighted composite where more discriminating items have a greater influence on trait estimation; DOT Total Correct, on the other hand, weights each item equally. The most straightforward indicator of cognitive processing speed is the time taken to answer all 48 DOT items (i.e., DOT Total Time).

Table 12 shows descriptive statistics for the three DOT variables discussed above for the total sample as well as for the SPs and SNFOs separately. As can be seen, there was little difference in DOT Total Correct or DOT IRT Score across the SP and SNFO groups. There were significant differences in processing speed, however, with SPs being moderately faster than SNFOs, with an effect size of approximately 0.3.

Table 12. DOT Performance Across SP and SNFO Groups

Program	N	DOT Total Correct		DOT IRT Score		DOT Total Time	
		Mean	SD	Mean	SD	Mean	SD
SNFO	89	35.39	8.75	.02	.88	231.85	92.91
SP	310	34.88	9.63	-.01	.98	203.09	82.27
Total	399	34.99	9.44	.00	.96	209.50	85.48

Table 13 shows correlations of the DOT scores with other potential predictors of training performance, such as ASTB scores and composites, as well as with education, past training, and simulator experience. DOT Total Correct and DOT IRT Score were modestly related (correlations of about .15) with previous flight simulator experience. These variables were more highly correlated with ASTB scores, with many correlations in the neighborhood of 0.3. Note that DOT Total Time exhibited the same pattern of correlations, albeit with the opposite sign and somewhat reduced magnitude.

Table 13. Correlations Between the DOT Scores and Other Predictors

	N	Mean	SD	DOT Total Correct	DOT IRT Score	DOT Total Time
aTraining	390	.23	.72	.043	.047	.045
Education	385	2.88	.59	.039	.035	.049
simExperience	399	.74	.83	.161**	.153**	-.132**
flightHours	391	.69	1.38	.001	.022	.005
ANI_RAW	332	.58	.53	.088	.095	-.070
MST_RAW	332	.34	.67	.184**	.189**	-.117*
RCT_RAW	332	.43	.53	.146**	.145**	-.101
SAT_Post2004	332	.76	.64	.309**	.327**	-.217**
MCT_Post2004	332	.50	.64	.307**	.316**	-.186**
AQR_Post2004	332	.55	.52	.326**	.338**	-.219**
PFAR_Post2004	332	.67	.50	.302**	.317**	-.206**
FOFAR_Post2004	332	.65	.53	.328**	.343**	-.232**
OAR_Post2004	332	.50	.62	.312**	.321**	-.193**

Note: ** indicates significance at $p < .01$

Table 14 presents correlations between three DOT scores (DOT Total Correct, DOT IRT Score and DOT Total Time) and training criteria (block grades and training composites) for the total sample as well as for the Student Pilots only. As can be seen in the table, DOT Total Correct and DOT IRT Score have many significant correlations with training performance. The correlations with the NSS are in the mid .20s, with slightly larger values for the SP sample. Again, the speed of cognitive processing (DOT Total Time) exhibits a similar pattern, but with the opposite sign and smaller magnitudes.

Table 14. Correlations Between the DOT Predictors and Navy Pilot Training Criteria

Training Block Name	Total Sample				Student Pilots (SPs)			
	N	DOT Total Correct	DOT IRT Score	DOT Total Time	N	DOT Total Correct	DOT IRT Score	DOT Total Time
C20	399	.122*	.147**	-.074	310	.152**	.179**	-.110
C40	86	-.027	.009	-.209	-	-	-	-
C41	374	.131*	.143**	-.072	292	.175**	.190**	-.079
C42	367	.103*	.103*	-.010	284	.125*	.131*	-.043
C43	270	.098	.119	-.051	270	.098	.119	-.051
C45	262	.174**	.174**	-.116	262	.174**	.174**	-.116
C46	246	.150*	.150*	-.104	246	.150*	.150*	-.104
C47	238	.101	.113	.073	238	.101	.113	.073
I20	387	.250**	.267**	-.121*	303	.274**	.287**	-.142*
I21	300	.219**	.236**	-.159**	300	.219**	.236**	-.159**
I22	207	.183**	.201**	-.209**	207	.183**	.201**	-.209**
I23	206	.116	.125	-.172*	206	.116	.125	-.172*

I24	194	.156*	.178*	-.219**	194	.156*	.178*	-.219**
I25	188	.214**	.218**	-.121	188	.214**	.218**	-.121
I40	380	.161**	.174**	-.091	298	.200**	.221**	-.137*
I41	282	.176**	.184**	-.105	200	.171*	.179*	-.180*
I42	233	.103	.098	-.106	183	.072	.071	-.100
I43	227	.102	.108	-.029	178	.071	.078	-.110
F40	225	.281**	.284**	-.147*	225	.281**	.284**	-.147*
F42	209	.204**	.189**	-.223**	209	.204**	.189**	-.223**
N40	183	.069	.082	-.088	183	.069	.082	-.088
N41	181	.011	.036	.004	181	.011	.036	.004
Contact_Simulation	399	.122*	.147**	-.074	310	.152**	.179**	-.110
Contact_AIRCRAFT	378	.152**	.168**	-.109*	292	.211**	.224**	-.085
Contact_ALL	399	.173**	.193**	-.096	310	.217**	.237**	-.090
Instruments_Simulation	387	.256**	.277**	-.151**	303	.287**	.303**	-.184**
Instruments_AIRCRAFT	380	.194**	.200**	-.122*	298	.226**	.236**	-.165**
Instruments_ALL	387	.261**	.276**	-.155**	303	.297**	.312**	-.197**
Instruments_BASIC	387	.250**	.268**	-.148**	303	.291**	.311**	-.185**
Instruments_RADIO	289	.187**	.198**	-.118*	207	.188**	.200**	-.212**
Instruments_NAVIGATION	244	.195**	.206**	-.126*	194	.182*	.198**	-.180*
Navigation_AIRCRAFT	183	.034	.058	-.039	183	.034	.058	-.039
Formation_AIRCRAFT	225	.275**	.269**	-.184**	225	.275**	.269**	-.184**
Navy Standard Score (NSS)	399	.232**	.251**	-.142**	310	.277**	.299**	-.158**

** indicates significance at $p < .01$

The final set of analyses involving the DOT scores concerns their operational use. Multiple regression analyses were run separately for DOT Total Correct and DOT IRT Score with DOT Total Time and the respective interaction terms. These analyses were then repeated using standardized variables because we believed z-scores would be easier to interpret. (Note: If one wants to use these regression weights for selection decisions, the raw scores must therefore be converted to z-scores prior to computing predicted criterion values.)

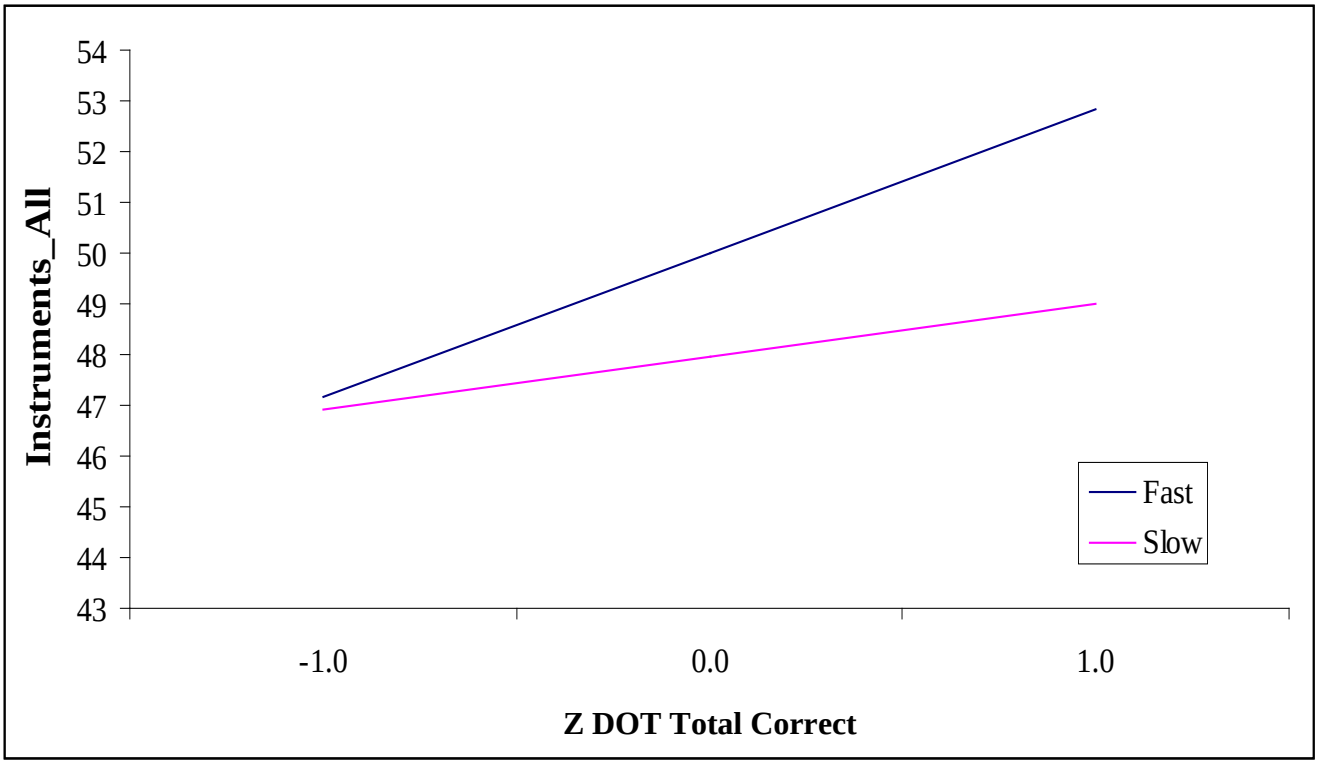
The results of the regression analyses involving standardized variables are presented in Table 15. All of the criterion variables except Navigation_Aircraft were predicted reasonably well, with multiple correlations around 0.3. The speed variable was significant in many of the models, as were the respective interaction terms. The interaction of the standardized DOT Total Time and DOT Total Correct variables is illustrated graphically in Figure 2. As expected, those who finished faster (lower times) had better grades than those who were slower, given the same total correct scores.

Table 15. Regression Results for Predicting Composite Training Criteria with DOT Scores

Criterion	Predictor Scores (z = Standardized)	Regression Coefficients B	SE	Sig.	R
Contact_All	(Constant)	48.977	.361	.000	.249
	z DOT Total Correct	1.173	.362	.001	
	z DOT Total Time	-.650	.365	.076	
	Interaction	-1.148	.335	.001	
Instruments_All	(Constant)	48.986	.396	.000	.308
	z DOT Total Correct	1.936	.397	.000	
	z DOT Total Time	-1.016	.397	.011	
	Interaction	-.898	.364	.014	
Navigation_Aircraft	(Constant)	50.355	.650	.000	.049
	z DOT Total Correct	.299	.739	.686	
	z DOT Total Time	-.315	.699	.653	
	Interaction	.107	.847	.900	
Formation_Aircraft	(Constant)	49.913	.573	.000	.326
	z DOT Total Correct	2.017	.597	.001	
	z DOT Total Time	-1.639	.632	.010	
	Interaction	-.874	.654	.183	
NSS Grades	(Constant)	48.766	.485	.000	.305
	z DOT Total Correct	2.125	.487	.000	
	z DOT Total Time	-1.257	.491	.011	
	Interaction	-1.591	.450	.000	
Contact_All	(Constant)	48.997	.360	.000	.256
	z DOT IRT Score	1.280	.361	.000	
	z DOT Total Time	-.583	.363	.109	
	Interaction	-1.058	.329	.001	
Instruments_All	(Constant)	48.998	.395	.000	.317
	z DOT IRT Score	2.014	.395	.000	
	z DOT Total Time	-.968	.395	.015	
	Interaction	-.837	.358	.020	
Navigation_Aircraft	(Constant)	50.379	.651	.000	.072
	z DOT IRT Score	.575	.722	.427	
	z DOT Total Time	-.258	.701	.714	

	Interaction	.327	.800	.683	
Formation_Aircraft	(Constant)	49.886	.574	.000	.323
	z DOT IRT Score	1.899	.598	.002	
	z DOT Total Time	-1.661	.634	.009	
	Interaction	-.943	.648	.147	
NSS Grades	(Constant)	48.789	.483	.000	.314
	z DOT IRT Score	2.262	.485	.000	
	z DOT Total Time	-1.170	.487	.017	
	Interaction	-1.487	.442	.001	

Figure 2. Interaction Between the Standardized Total Correct Scores and the Standardized Total Response Time When Predicting the Instruments All Training Composite



In summary, the DOT subtest was found to predict many criterion variables well. Significant effects were found for the standardized predictor scores involving total response time, total correct, and the interaction of the two, so it is recommended that regression-based composites for selection purposes include an interaction term. In addition, it was found that the 3PLM fit the DOT data very well and it can be applied with confidence in future investigations involving differential item or test functioning analyses.

DICHOTIC LISTENING TEST (DLT): SCORING STRATEGIES AND VALIDITIES

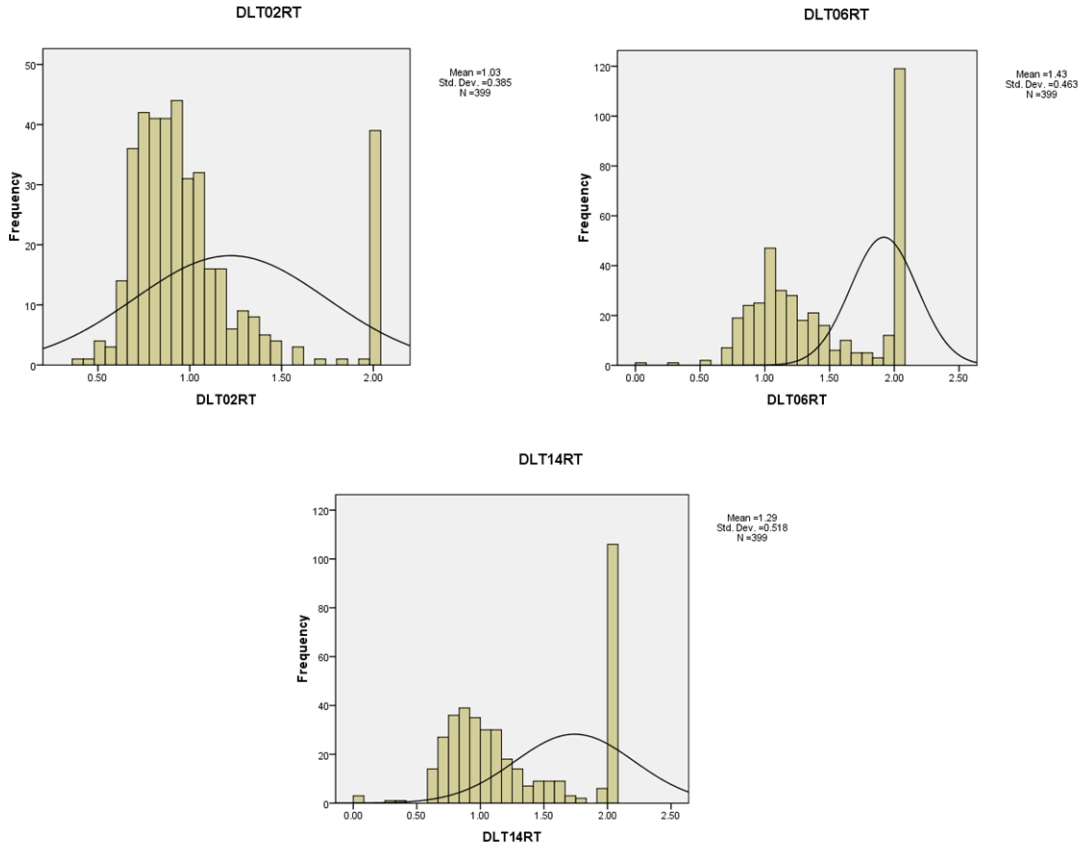
During the DLT, an examinee wearing headphones is presented with a different series of letters and numbers in each ear. The examinee is instructed to monitor a "target" ear and press the trigger on a joystick when an even number is presented to the target ear and to press the thumb button (RDR Cursor Button) on a throttle when an odd number is presented to the target ear. The examinees are instructed to ignore numbers when presented to a non-target ear in addition to ignoring letters at all times.

The DLT involves four sets of trials, each lasting 30 seconds and involving four numbers in the target ear. Thus, 16 targets are presented to each examinee during the course of a test. When an examinee detects a number in the target ear, he/she has 2000ms to perform the designated action. Only the first response action following the stimulus presentation is recorded. If the action is correct, then a score of "1" is recorded; otherwise, a score of zero is assigned for that "item." The current PBM software does not track false positive responses, so the minimum number correct score is 0 and the maximum number correct score is 16.

Similar to the DOT, the DLT measures a combination of abilities/skills: auditory recognition (i.e., aural comprehension), cognitive processing speed via response time, and psychomotor dexterity, which comes into play because the examinee must manipulate a button on a throttle or a trigger on a joystick in response to a perceived target. However, unlike the DOT, where the number of correct responses and response times are recorded independently for scoring purposes, the DLT requires an examinee to respond to a target within 2000ms or an item score of 0 is recorded, and there is currently no way to differentiate auditory recognition errors from incorrect motor responses or slow reaction times. Additionally, the current DLT captures the first response given within the 2000ms reaction time window after stimulus presentation. It does not capture any subsequent responses given prior to presentation of the next stimulus, which is unfortunate. We did not analyze the DLT data at an item-level, but we proceeded with an examination of subtest scores in relation to criterion variables. For illustration, the reaction time distributions for three DLT items (DLT02, DLT06 and DLT14) are presented below in Figure 3.

The histograms in the figure indicate that response times were positively skewed, with an unusually high peak for the 2000ms category. These values represent omitted responses, where the examinee failed to press either button during the 2000ms data capture window, as well as any response latencies of exactly 2000ms, although latencies near this value appear to be extremely rare, as depicted in Figure 3. It seems unlikely that any of the 2000ms data points represent actual responses. Future software development efforts can hopefully differentiate examinee errors and detect responses made in the absence of an aural stimulus.

Figure 3. Response Time Distributions for Three Illustrative DLT Items



Note: RT values depicted at 2000ms in this figure represent both RTs at 2000ms as well as omitted responses.

Analysis of DLT Scale Scores

Table 16 shows descriptive statistics for the DLT Total Correct scores for the total sample as well as for the SPs and SNFOs separately. As can be seen, the difference between the two groups was small with an effect size less than 0.1. Therefore, although SNFOs had a slightly higher mean DLT score, the samples were combined and analyzed together. Figure 4 shows the frequency distribution for DLT Total Correct in the total sample.

Table 16. DLT Performance Across SP and SNFO Groups

Program	N	DLT Total Correct	
		Mean	SD
SNFO	89	10.60	5.35
SP	310	11.22	4.93
Total	399	11.08	5.02

Figure 4. Frequency Distribution for DLT Total Correct in the Total Sample

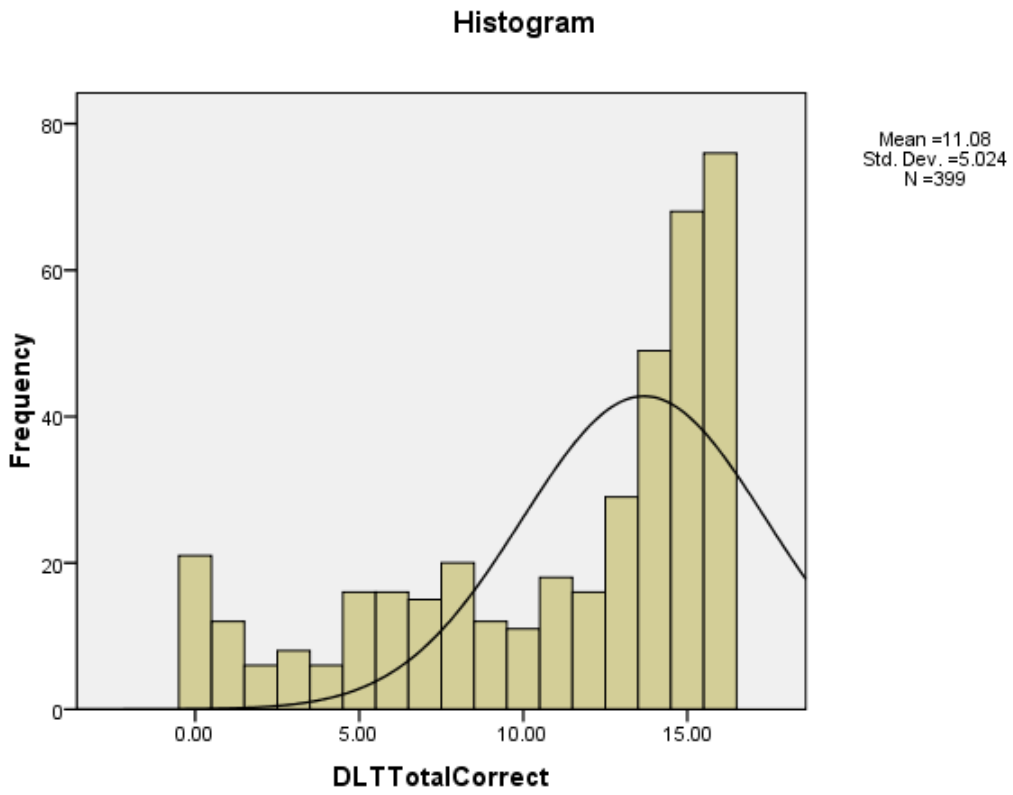


Table 17 shows the correlations of DLT scores with other potential predictors of training performance such as ASTB scores and composites as well as education, past training, and simulator experience. The correlations are generally small, suggesting that multicollinearity may not be a problem in multiple regression analyses.

Table 17. Correlations Between the DLT Scores and Other Predictors

	N	Mean	SD	DLT Total Correct
aTraining	390	.23	.72	.001
Education	385	2.88	.59	.112*
simExperience	399	.74	.83	.052
flightHours	391	.69	1.38	-.043
ANI_RAW	332	.58	.53	-.030
MST_RAW	332	.34	.67	.216**
RCT_RAW	332	.43	.53	.132*
SAT_Post2004	332	.76	.64	.091
MCT_Post2004	332	.50	.64	.194**
AQR_Post2004	332	.55	.52	.187**
PFAR_Post2004	332	.67	.50	.109*
FOFAR_Post2004	332	.65	.53	.183**
OAR_Post2004	332	.50	.62	.239**

Table 18 presents correlations between the DLT scores and training criteria (block grades and training composites) for the total sample as well as for the student pilots only. As can be seen in the table, these correlations are substantially lower than the correlations of the DOT with the training criterion measures. In fact, the DLT correlated only 0.11 with the NSS.

Table 18. Correlations Between the DLT Scores and Navy Pilot Training Criteria

Training Block Name	Total Sample		Students Pilots (SPs)	
	N	DLT Total Correct	N	DLT Total Correct
C20	399	.153**	310	.124*
C40	86	.047	-	-
C41	374	-.014	292	.060
C42	367	.050	284	.068
C43	270	.115	270	.115
C45	262	.009	262	.009
C46	246	.005	246	.005
C47	238	.014	238	.014
I20	387	.125*	303	.117*
I21	300	.093	300	.093
I22	207	.074	207	.074
I23	206	.029	206	.029
I24	194	.067	194	.067
I25	188	.071	188	.071
I40	380	.081	298	.097
I41	282	.035	200	.014
I42	233	-.009	183	.004
I43	227	-.022	178	-.035
F40	225	.103	225	.103
F42	209	.097	209	.097
N40	183	-.029	183	-.029
N41	181	.061	181	.061
Contact_Simulation	399	.153**	310	.124*
Contact_AIRCRAFT	378	.044	292	.061
Contact_ALL	399	.108*	310	.107
Instruments_Simulation	387	.111*	303	.096
Instruments_AIRCRAFT	380	.086	298	.103
Instruments_ALL	387	.102*	303	.103
Instruments_BASIC	387	.120*	303	.123*
Instruments_RADIO	289	.051	207	.040

Instruments_NAVIGATION	244	.060	194	.084
Navigation_AIRCRAFT	183	.058	183	.058
Formation_AIRCRAFT	225	.110	225	.110
Navy Standard Score (NSS)	399	.113*	310	.118*

VERTICAL TRACKING TEST (VTT):
SCORING STRATEGIES AND VALIDITIES

The main source of information about an examinee's performance on the VTT comes from Euclidian distances between the crosshairs and the airplane target, which are recorded during the test. The distance is checked every 35ms and, if the examinee is "on target" (i.e., the distance is zero), a counter is incremented until it reaches 40, initiating a random shift in the aircraft's direction, called a "redirect". The total number of redirects for the subtest is an indication of how many times the person was on target across these 35ms intervals, with higher numbers indicating more time spent on target.

In addition to the total number of redirects, the PBM records and stores Euclidian distances between the crosshairs and the airplane target every 400ms during the test duration. There are a total of 147 distances saved for the VTT; 50 captured during the first 20 seconds while the airplane's speed is slow, 50 during the next 20 seconds when the airplane's speed increases (i.e., "medium") and the final 47 captured during the final 20 seconds when the airplane's speed is fast. No distance data are captured for the final 1.2 seconds of the VTT. The PBM program computes the average distance between the crosshairs and the airplane target across these 147 time points, as well as the number of times the examinee was on target during slow, medium, and fast 20 second intervals. We have computed an additional score, the Total On Target, which is the sum of on-target counts for these three 20 second speed intervals.

Note that although all these subtest-level VTT scores are interrelated (all based on a similar source), their validities were explored separately in an effort to identify the most robust way to capture examinee performance on the VTT subtest.

In addition to these summary indices of VTT performance, an "item-level" index was developed using the 147 distance values, in order to permit polytomous IRT modeling, and therefore differential item and test functioning analyses. These are desirable since they permit evaluation of potential test bias in the absence of criterion data.

The item-level VTT performance index was developed using the 147 distance values captured every 400ms, dichotomized to represent on-target status (1 = on-target with distance at zero pixels; 0 = off-target with pixel distance greater than zero). The first 41 400ms intervals within each speed variation period (slow, medium, and fast) were condensed as follows to yield 3 polytomous items for each speed variation, for a total of 9 polytomous items capturing VTT on-target performance.

Within each speed variation, three periods of 13 adjacent dichotomized on-target measures, each representing 5.2 seconds of data, were summed to yield a score between 0 and 13. The first thirteen dichotomous distance measures were summed to yield a continuous performance index across this 5.2 second interval with possible values between 0 and 13. One 400ms interval was skipped to reduce dependency between adjacent dichotomous distance measures, and then dichotomized performance across the next 5.2 second (13 400ms interval) period was summed.

One more 400 ms interval was skipped, and then performance was summed across a third 5.2 second interval.

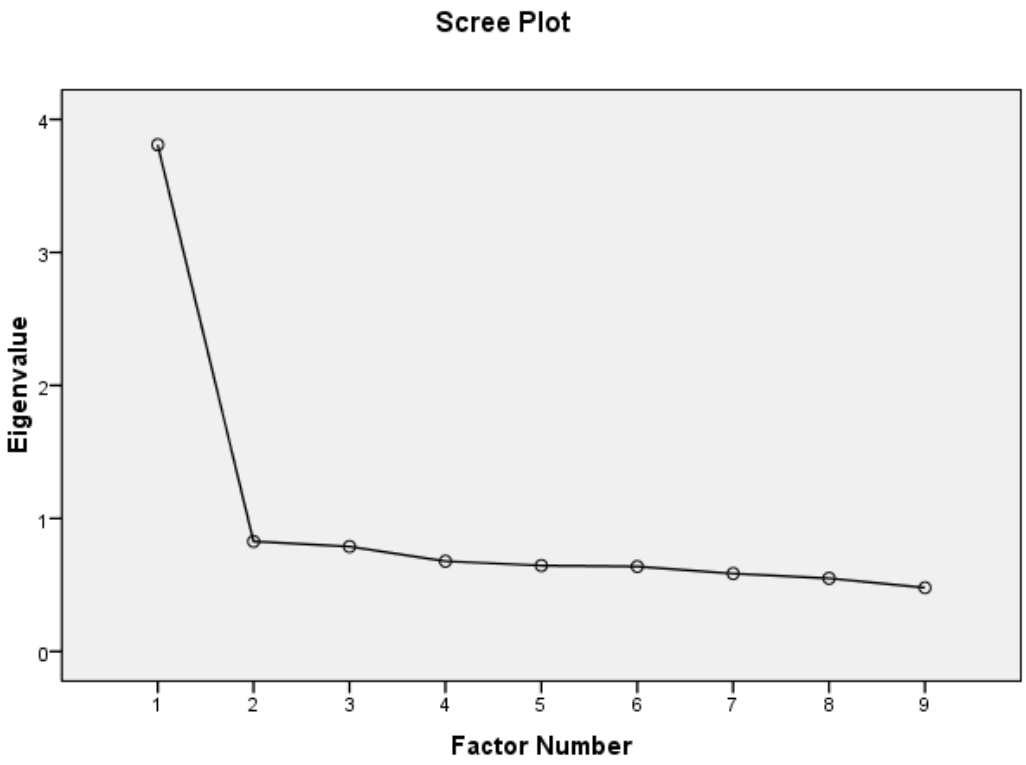
This process was repeated using the first 41 400ms intervals (13, 1 skipped, 13, 1 skipped, 13, and the remaining 6 or 9 400ms intervals skipped) from each 20 second speed-specific period to yield 9 scores ranging from 0 to 13. To create polytomous items with five response options, these 9 scores were collapsed into 5 categories according to the following scheme: 0-2 = 0; 3-4 = 1; 5-6 = 2; 7-8 = 3, 9-13 = 4.

As stated above, the main advantage of converting continuous data into categorical data is that polytomous IRT models could be fit to VTT data, making it possible to conduct differential item and test functioning analyses. The disadvantage is that there is some loss of psychometric information. Here, this was not a particular concern because the nine polytomous response variables seemed likely to capture nearly all of the available information.

Item-Level CTT and IRT Analyses and Results for the VTT

Because the response data were scored polytomously with each response score indicating a graded increment in examinee's ability to track a moving airplane, we chose Samejima's graded response model (SGRM; Samejima, 1969) as the basis for item analysis. To check the unidimensionality assumption of SGRM, we ran a principal component analysis (PCA) on the inter-item correlations and plotted the eigenvalues to produce the scree plot shown in Figure 5. As can be seen in the figure, the ratio of first to second eigenvalues exceeded 3.0 and there was a smooth tail formed by the second and subsequent eigenvalues, signifying that the response data were sufficiently unidimensional for the application of SGRM (Drasgow & Parsons, 1983; Lord, 1980).

Figure 5. Scree Plot for the Principal Component Analysis of the 9 VTT Items



IRT Calibration of the 9 VTT Items

SGRM was used to analyze the VTT data because of the ordered polytomous coding of the distance values (Samejima, 1969). For SGRM, the probability of observing a particular response category depends on the discriminating power of the item, the extremity parameter for that category, and the value of the latent trait (theta) representing examinee ability. The equation for computing SGR category response probabilities is:

$$P_i(v_i = j \mid \theta) = \frac{1}{1+\exp[-1.7a_i(\theta - b_{i,j})]} - \frac{1}{1+\exp[-1.7a_i(\theta - b_{i,j+1})]},$$

where v_i denotes a scored response to item i; j is an index for response categories ($j = 1, \ldots, J$, where J refers to the number of categories for the item); a_i is the item discrimination parameter, which is assumed to be the same for all categories associated with an item; and b is the extremity or threshold parameter that varies from category to category, given the constraints $b_{j-1} < b_j < b_{j+1}$ and b_J is taken to be $+\infty$.

The discrimination parameter for SGRM can be interpreted in the same way as the discrimination parameter for the 3PLM. The option extremity parameters, b_1, b_2, … b_{J-1}, for a J option polytomous item can be interpreted as follows: The first extremity parameter, b_1, corresponds to the point along the latent trait continuum where respondents have a 50% chance of obtaining a score of 0, the second extremity parameter, b_2, corresponds to the point along the latent trait continuum where respondents have a 50% chance of obtaining a score of 1 or less, etc.

The MULTILOG computer program (Thissen, 1991) was used to estimate SGRM item parameters, and the response data were scored using the MODFIT-Z 2.0 computer program (Stark, 2007). The MULTILOG command file is shown below.

```
PBM VTT subtest
graded model
>PROBLEM RANDOM,
      INDIVIDUAL,
      DATA = 'VTT399.DAT',
      NITEMS = 9,
      NGROUPS = 1,
      NEXAMINEES = 399,
      NCHARS = 5;
>TEST ALL,
    GRADED,
    NC = (5(0)9);
>ESTIMATE NCYCLES=200 , ITERATIONS=50;
>TGROUPS  NUMBER=31,  QP=(-4.5(0.3)4.5);
>SAVE;
>END ;
5
01234
111111111
222222222
333333333
444444444
555555555
(5a1,9a1)
```

As in the DOT analysis, the fit of SGRM to the VTT data was examined using fit plots and chi-square statistics computed using MODFIT-Z 2.0 (Stark, 2007). Overall the fit plots indicated that SGRM fit the VTT response data well, and this finding was confirmed by the chi-square analyses, which yielded means of 0.00, 0.72, and 0.52 for singlets, doublets, and triplets respectively. The frequency distribution for adjusted chi-square values is shown in Table 19.

Table 19. Chi-Square Model-Data Fit Statistics for Items Created from the VTT Data

FREQUENCY TABLE OF ADJUSTED (N=3000) CHISQUARE/DF RATIOS

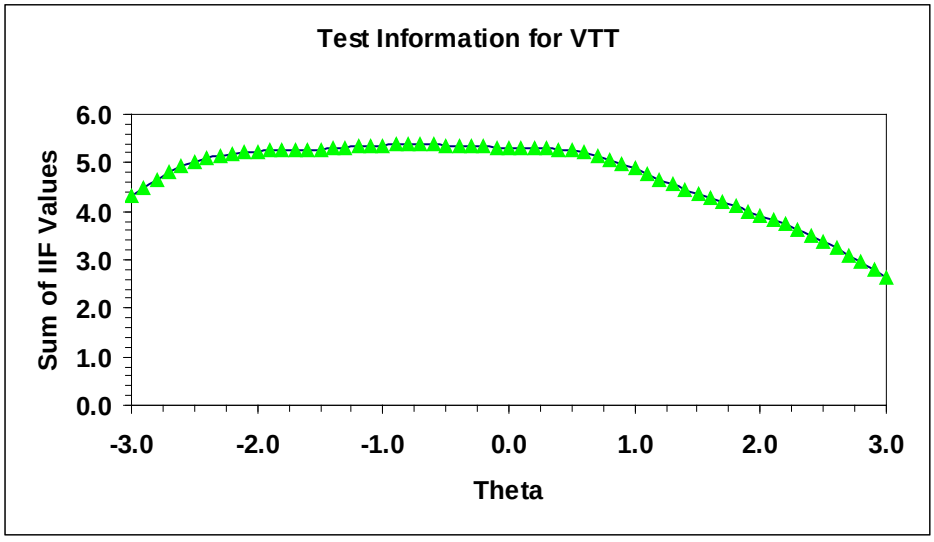
	<1	1<2	2<3	3<4	4<5	5<7	>7	Mean	SD
Singlets	9	0	0	0	0	0	0	0.00	0.00
Doublets	30	1	1	1	1	2	0	0.72	1.77
Triplets	69	5	7	2	1	0	0	0.52	0.99

Table 20 presents CTT statistics and IRT parameter estimates for the 9 VTT items. Shown are the item means, standard deviations (SD), corrected item-total correlations (CITC), SGRM item discrimination (a), and extremity parameters (b_1, b_2, b_3, *and* b_4). Note that all of the corrected item-total correlations are large (>.4), and that the IRT *a* parameter estimates are all fairly high in magnitude (>.8), with just a few exceptions (Note that these values do not include the 1.7 scaling factor). Moreover, the wide range of IRT *b* parameter estimates suggests that the test provides good measurement across a broad range of examinee ability. This is illustrated by the test information function shown in Figure 6.

Table 20. CTT and IRT Statistics for the 9 VTT Items

VTT Item Name	Polytomous Responses			SGRM Parameters				
	Mean	SD	CITC	*a*	b_1	b_2	b_3	b_4
VTT_slow1p	2.11	1.10	.54	0.85	-2.20	-0.77	.44	2.06
VTT_slow2p	2.86	1.06	.61	1.06	-2.66	-1.74	-.62	.66
VTT_slow3p	3.07	.93	.55	0.90	-3.86	-2.30	-1.12	.50
VTT_med1p	2.20	.97	.52	0.81	-2.97	-1.10	.41	2.25
VTT_med2p	2.10	1.05	.56	0.90	-2.34	-0.76	.49	2.14
VTT_med3p	2.24	1.03	.55	0.85	-2.51	-1.07	.30	1.98
VTT_fast1p	1.94	1.01	.49	0.75	-2.44	-0.66	.92	2.74
VTT_fast2p	1.97	.96	.44	0.62	-3.13	-0.87	.94	3.43
VTT_fast3p	2.00	1.01	.51	0.76	-2.61	-0.73	.78	2.57

Figure 6. Test Information Function for the 9 VTT Items



VTT Scale Scores

The total number of redirect (VTT Redirects), the average distances between the crosshairs and the airplane target during the test (VTT Average Distance), the total number of on-target responses (VTT Total On Target), and the IRT VTT Score are all indicators of examinee ability. Although they are highly correlated, each taps a somewhat different aspect of examinee performance. The VTT Average Distance is negatively related to the rest of the VTT scores, because a large score reflects *poor* performance and large scores for all of the other measures indicate *good* performance.

Table 21 shows descriptive statistics for the four VTT variables discussed above for the total sample as well as for the SPs and SNFOs separately. As can be seen, SPs outperformed the SNFOs on all measures, with the mean score typically better by about a third of the total standard deviation.

Table 21. VTT Performance Across SP and SNFO Student Groups

Program	N	VTT Redirects		VTT Average Distance		VTT Total On Target		VTT IRT Score	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
SNFO	89	16.19	3.50	28.16	10.95	76.76	16.19	-0.28	0.96
SP	310	17.43	3.11	24.98	9.50	82.64	14.30	0.09	0.87
Total	399	17.16	3.24	25.69	9.92	81.33	14.92	0.01	0.90

Table 22 shows correlations of VTT scores with other potential predictors of training performance, such as ASTB scores and composites, as well as with education, past training, and simulator experience. As expected, VTT scores were modestly related to simulator experience (i.e., simExperience; correlations of about 0.22), as well as with ASTB scores, with the absolute value of many correlations being around 0.20. Note also that VTT Average Distance tended to have the smallest correlations with other predictors.

Table 22. Correlations Between the VTT Scores and Other Predictors

	N	Mean	Std. Deviation	VTT Redirects	VTT Average Distance	VTT Total On Target	VTT IRT Score
aTraining	390	.23	.72	.054	-.056	.055	.039
Education	385	2.88	.59	.074	-.035	.072	.064
simExperience	399	.74	.83	.220**	-.149**	.224**	.227**
flightHours	391	.69	1.38	.039	-.037	.042	.043
ANI_RAW	332	.58	.53	.101	-.112*	.116*	.119*
MST_RAW	332	.34	.67	.111*	-.071	.101	.082
RCT_RAW	332	.43	.53	.050	-.044	.040	.052
SAT_Post2004	332	.76	.64	.187**	-.179**	.180**	.190**
MCT_Post2004	332	.50	.64	.225**	-.159**	.205**	.197**

AQR_Post2004	332	.55	.52	.235**	-.190**	.225**	.218**
PFAR_Post2004	332	.67	.50	.228**	-.203**	.226**	.227**
FOFAR_Post2004	332	.65	.53	.210**	-.184**	.204**	.200**
OAR_Post2004	332	.50	.62	.215**	-.150**	.195**	.183**

Table 23 presents correlations between four VTT scores and training criteria (block grades and training composites) for the total sample as well as for the student pilots only. As can be seen in the table, some of the correlations are large enough to have practical importance. For example, VTT Redirects and VTT Total On Target correlate about .24 with Formation_AIRCRAFT, and they correlate .21 with NSS for the SPs. Because there was only modest multicollinearity of the VTT scores with alternate predictors, it seemed likely that they would provide incremental validity.

Table 23. Correlations Between the VTT Predictors and Navy Pilot Training Criteria

Training Block Name	Total Sample					Students Pilots (SPs)				
	N	VTT Redirects	VTT Average Distance	VTT Total On Target	VTT IRT Score	N	VTT Redirects	VTT Average Distance	VTT Total On Target	VTT IRT Score
C20	399	.074	-.046	.073	.061	310	.104	-.077	.105	.098
C40	86	.111	-.068	.110	.067	-	-	-	-	-
C41	374	.082	-.074	.090	.089	292	.175**	-.143*	.178**	.178**
C42	367	.081	-.066	.082	.095	284	.115	-.105	.121*	.145*
C43	270	.036	-.030	.024	.020	270	.036	-.030	.024	.020
C45	262	.099	-.061	.097	.113	262	.099	-.061	.097	.113
C46	246	.093	-.077	.096	.075	246	.093	-.077	.096	.075
C47	238	.131*	-.072	.106	.126	238	.131*	-.072	.106	.126
I20	387	.182**	-.118*	.173**	.164**	303	.195**	-.133*	.186**	.196**
I21	300	.216**	-.158**	.202**	.209**	300	.216**	-.158**	.202**	.209**
I22	207	.113	-.070	.127	.128	207	.113	-.070	.127	.128
I23	206	.103	-.054	.091	.077	206	.103	-.054	.091	.077
I24	194	.196**	-.133	.175*	.170*	194	.196**	-.133	.175*	.170*
I25	188	.114	-.055	.085	.100	188	.114	-.055	.085	.100
I40	380	.072	-.073	.078	.091	298	.117*	-.119*	.122*	.134*
I41	282	.174**	-.143*	.172**	.145*	200	.175*	-.136	.173*	.159*
I42	233	.121	-.092	.102	.080	183	.138	-.106	.116	.097
I43	227	.021	-.038	.015	.025	178	.012	-.033	.009	.016
F40	225	.209**	-.148*	.214**	.207**	225	.209**	-.148*	.214**	.207**
F42	209	.232**	-.207**	.237**	.214**	209	.232**	-.207**	.237**	.214**
N40	183	.111	-.140	.111	.115	183	.111	-.140	.111	.115
N41	181	.032	-.068	.045	.047	181	.032	-.068	.045	.047

Contact_Simulation	399	.074	-.046	.073	.061	310	.104	-.077	.105	.098
Contact_AIRCRAFT	378	.141**	-.108*	.140**	.136**	292	.174**	-.142*	.172**	.182**
Contact_ALL	399	.123*	-.088	.122*	.109*	310	.149**	-.115*	.149**	.147**
Instruments_Simulation	387	.191**	-.127*	.181**	.163**	303	.206**	-.146*	.197**	.197**
Instruments_AIRCRAFT	380	.064	-.058	.066	.066	298	.113	-.108	.115*	.115*
Instruments_ALL	387	.151**	-.106*	.146**	.138**	303	.194**	-.148**	.188**	.188**
Instruments_BASIC	387	.171**	-.121*	.167**	.165**	303	.207**	-.159**	.201**	.212**
Instruments_RADIO	289	.144*	-.111	.148*	.125*	207	.136	-.092	.142*	.135
Instruments_NAVIGATION	244	.147*	-.104	.119	.115	194	.173*	-.119	.140	.137
Navigation_AIRCRAFT	183	.103	-.130	.108	.112	183	.103	-.130	.108	.112
Formation_AIRCRAFT	225	.238**	-.187**	.244**	.227**	225	.238**	-.187**	.244**	.227**
Navy Standard Score (NSS)	399	.162**	-.117*	.159**	.147**	310	.211**	-.160**	.206**	.202**

In summary, the VTT appears to reliably measure a largely unidimensional tracking ability. This ability correlates meaningfully with some aspects of training performance and therefore the VTT appears to be a strong candidate for a pilot training selection battery.

AIRPLANE TRACKING TEST (ATT): SCORING STRATEGIES AND VALIDITIES

Similarly to the VTT subtest, the main source of information about an examinee's performance on the ATT comes from Euclidian distances between the crosshairs and the airplane target, which are recorded during the test. The distance is checked every 35ms and, if the examinee is "on target" (i.e., the distance is zero), a counter is incremented until it reaches 30, initiating a redirect just as was the case for the VTT. The reason for this lower threshold for redirect initiation in the ATT is the increased relative difficulty of placing the cursor directly over a target moving in two dimensions. The total number of such directional shifts (redirects) for the whole is an indication of how many times the person was on target during the test, with higher numbers indicating more time spent on target.

In addition to the total number of redirects, the PBM records and stores Euclidian distances between the crosshairs and the airplane target every 400ms during the test. As in the VTT subtest, there are a total of 147 distances saved for the ATT: 50 captured during the first 20 seconds while the airplane's speed is slow, 50 during the next 20 seconds when the airplane's speed increases (i.e., "medium") and the final 47 captured during the final 20 seconds when the airplane's speed is fast. No distance data are captured for the final 1.2 seconds of the ATT. The PBM program computes the average distance between the crosshairs and the airplane target across these 147 time points, as well as the number of times the examinee was on target during slow, medium and fast 20 second intervals. We computed an additional score, the Total On Target, which is the sum of on-target counts for the three airplane speeds.

Note that although all these subtest-level ATT scores are interrelated (all based on a similar source), their validities were explored separately in an effort to identify the most robust way to capture examinee performance on ATT.

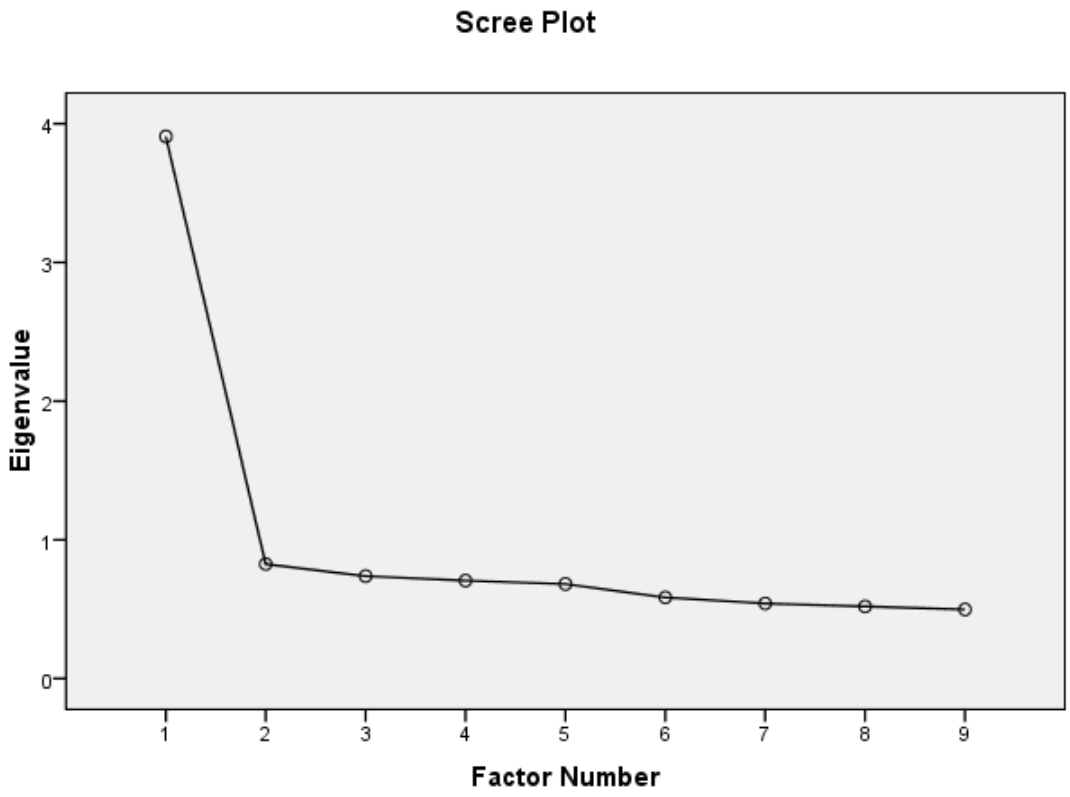
In addition to these summary indices of ATT performance, an "item-level" index was developed using the same procedure described for the VTT in order to permit polytomous IRT modeling. Similar to the VTT, this yielded 9 5-option polytomous items representing examinees' ability to keep the crosshairs on target over most of this 60-second subtest.

It should be noted that very few examinees had scores higher than 8 on any of the 13 intervals used to produce polytomous items due to the difficulty of the ATT subtest.

Item-Level CTT and IRT Analyses and Results for the ATT

Because the response data were scored polytomously with category codes of higher magnitude indicating better performance on the two-dimensional airplane tracking task, SGRM (Samejima, 1969) for ordered polytomous responses was chosen for IRT analysis. To verify that the response data were sufficiently unidimensional, we conducted a principal component analysis (PCA) of the inter-item correlations. The resulting scree plot is shown in Figure 7 below. As can be seen, the data exhibited a strong first factor with the ratio of first to second eigenvalues exceeding the 3.0 rule of thumb suggested for analysis with unidimensional IRT models (Drasgow & Parsons, 1983; Lord, 1980).

Figure 7. Scree Plot for the Principal Component Analysis of the 9 ATT Items



IRT Calibration of the 9 ATT Items

Because the ATT data were coded such that the responses to each item fell into one of five ordered categories, there were five parameters to estimate for each item: one discrimination parameter, a, and four extremity parameters, b_1, b_2, b_3, and b_4 (see the description of SGRM in the VTT section of this report for details).

The MULTILOG computer program (Thissen, 1991) was used to estimate SGRM item parameters, and the response data were scored using the MODFIT-Z 2.0 computer program (Stark, 2007). The MULTILOG command file is shown below.

```
PBM ATT subtest
graded model
>PROBLEM RANDOM,
      INDIVIDUAL,
      DATA = 'ATT399.DAT',
      NITEMS = 9,
      NGROUPS = 1,
      NEXAMINEES = 399,
      NCHARS = 5;
>TEST ALL,
    GRADED,
    NC = (5(0)9);
>ESTIMATE NCYCLES=200 , ITERATIONS=50;
>TGROUPS  NUMBER=31,  QP=(-4.5(0.3)4.5);
>SAVE;
>END ;
5
01234
111111111
222222222
333333333
444444444
555555555
(5a1,9a1)
```

As in the VTT analysis, the fit of SGRM to the ATT data was examined using fit plots and chi-square statistics computed via MODFIT-Z. Overall the fit plots indicated that SGRM fit the ATT response data well, and this finding was supported by the chi-square analyses, which yielded means of 0.00, 0.11, and 0.28 for singlets, doublets, and triplets respectively. The frequency distribution for the adjusted chi-square values is shown in Table 24.

Table 24. Chi-Square Model-Data Fit Statistics for Items Created from the ATT Data

FREQUENCY TABLE OF ADJUSTED (N=3000) CHISQUARE/DF RATIOS

	<1	1<2	2<3	3<4	4<5	5<7	>7	Mean	SD
Singlets	9	0	0	0	0	0	0	0.00	0.00
Doublets	34	2	0	0	0	0	0	0.11	0.31
Triplets	75	5	4	0	0	0	0	0.28	0.68

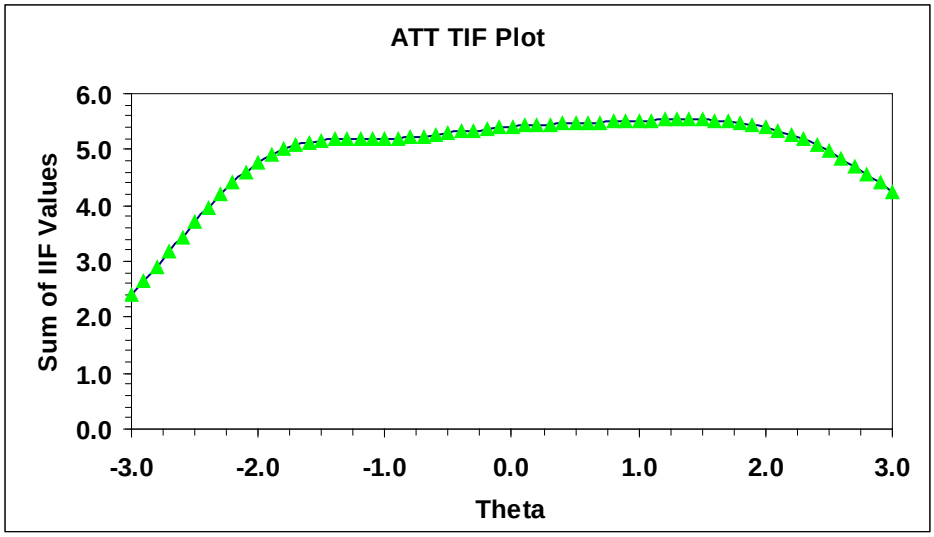
Table 25 presents CTT statistics and IRT parameter estimates for the 9 ATT items. Shown are the item means, standard deviations (SD), corrected item-total correlations (CITC), SGRM item discrimination (a), and extremity parameters (b_1, b_2, b_3, and b_4). Note that all of the corrected item-total correlations are large, with several approaching .6, and the IRT *a* parameter estimates are correspondingly high with many for the slow and medium parts of the test in the 0.9 to 1.0 range (These values do not include the 1.7 scaling factor). These results indicate that the ATT

items are effective at discriminating between examinees of different levels of ability. Moreover, the wide range of IRT *b* parameter estimates suggests that the test provides good measurement precision across a broad range of examinee ability. This is illustrated by the test information function shown in Figure 8.

Table 25. CTT and IRT Statistics for the 9 ATT Items

ATT Item Name	Polytomous Responses			SGRM Parameters				
	Mean	SD	CITC	*a*	b_1	b_2	b_3	b_4
ATT_slow1p	1.65	1.06	.58	0.96	-1.58	-0.14	1.28	2.38
ATT_slow2p	1.94	1.08	.58	0.97	-2.15	-0.49	.82	1.91
ATT_slow3p	2.01	1.10	.59	0.92	-2.03	-0.65	.66	1.99
ATT_med1p	1.54	.96	.56	0.88	-1.69	0.02	1.58	3.04
ATT_med2p	1.61	1.02	.49	0.71	-1.90	-0.09	1.50	3.24
ATT_med3p	1.67	1.01	.58	0.93	-1.76	-0.18	1.21	2.65
ATT_fast1p	1.37	.94	.51	0.75	-1.63	0.43	1.96	3.69
ATT_fast2p	1.34	.92	.49	0.74	-1.56	0.41	2.15	4.01
ATT_fast3p	1.32	.88	.53	0.79	-1.64	0.53	2.18	3.79

Figure 8. Test Information Function for the 9 ATT Items



ATT Scale Scores

The total number of redirects (ATT Redirects), the average distance between the crosshairs and the airplane target during the test (ATT Average Distance), the total number of on-target responses (ATT Total On Target), and the IRT ATT Score are all indicators of examinee ability. Although they are highly correlated, each taps a somewhat different aspect of examinee performance. The ATT Average Distance is negatively related to the rest of the ATT scores,

because a large score in this case reflects *poor* performance, whereas large scores on the other measures indicate *good* performance.

Table 26 shows descriptive statistics for the four ATT variables discussed above in the total sample and across SP and SNFO student groups. As can be seen, SPs outperformed the SNFOs on all measures, with the mean scores better by about one third to one half of the total group standard deviation.

Table 26. ATT Performance Across SP and SNFO Student Groups

Program	N	ATT Redirects		ATT Average Distance		ATT Total On Target		ATT IRT Score	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
SNFO	89	7.46	4.37	81.39	33.34	27.89	14.86	-0.27	0.99
SP	310	9.19	4.21	68.97	24.35	33.45	14.39	0.07	0.88
Total	399	8.80	4.30	71.74	27.08	32.21	14.66	0.00	0.92

Table 27 shows correlations of ATT scores with other potential predictors of training performance, such as ASTB scores and composites, as well as with education, past training, and simulator experience. In contrast with the DOT and VTT, ATT scores showed distinctly higher correlations with simExperience (.31 or higher versus .23 or lower in the previous cases), reflecting the more challenging nature of the ATT and improvements in skill that likely derive from practice, but the correlations with flightHours were still relatively small. Moreover, ATT Average Distance showed negative correlations with the other ATT scores, as expected, and it possessed the smallest correlations with the other predictors.

Table 27. Correlations Between the ATT Scores and Other Predictors

	N	Mean	Std. Deviation	ATT Redirects	ATT Average Distance	ATT Total On Target	ATT IRT Score
aTraining	390	.23	.72	.080	-.082	.072	.055
Education	385	2.88	.59	.018	.000	.028	.025
simExperience	399	.74	.83	.372**	-.310**	.363**	.332**
flightHours	391	.69	1.38	.140**	-.110*	.134**	.098
ANI_RAW	332	.58	.53	.256**	-.204**	.239**	.209**
MST_RAW	332	.34	.67	.124*	-.088	.134*	.119*
RCT_RAW	332	.43	.53	.064	-.064	.061	.055
SAT_Post2004	332	.76	.64	.210**	-.196**	.233**	.212**
MCT_Post2004	332	.50	.64	.220**	-.176**	.231**	.202**
AQR_Post2004	332	.55	.52	.310**	-.250**	.317**	.278**
PFAR_Post2004	332	.67	.50	.336**	-.277**	.338**	.299**
FOFAR_Post2004	332	.65	.53	.287**	-.241**	.301**	.267**
OAR_Post2004	332	.50	.62	.217**	-.170**	.229**	.201**

Table 28 presents correlations between the four ATT scores and training criteria (block grades and training composites) for the total sample as well as for the student pilots only. As can be seen in the table, some of the correlations are large enough to have practical importance. For example, ATT Redirects and ATT Total On Target correlated as high as .29 (in magnitude) with individual instruments criteria (e.g., I20), .26 for Formation_AIRCRAFT, and .24 for Instruments_ALL in the total samples. The respective results for the SP group were slightly higher with magnitudes of .32, .27, and .26.

Table 28. Correlations Between the ATT Predictors and Navy Pilot Training Criteria

Training Block Name	Total Sample					Students Pilots (SPs)				
	N	ATT Redirects	ATT Average Distance	ATT Total On Target	ATT IRT Score	N	ATT Redirects	ATT Average Distance	ATT Total On Target	ATT IRT Score
C20	399	.057	-.029	.065	.035	310	.078	-.077	.087	.060
C40	86	.082	-.023	.077	.063	-	-	-	-	-
C41	374	.117*	-.077	.120*	.100	292	.181**	-.159**	.185**	.159**
C42	367	.071	-.082	.088	.075	284	.140*	-.144*	.150*	.129*
C43	270	.102	-.122*	.101	.084	270	.102	-.122*	.101	.084
C45	262	.177**	-.145*	.180**	.171**	262	.177**	-.145*	.180**	.171**
C46	246	.143*	-.196**	.140*	.133*	246	.143*	-.196**	.140*	.133*
C47	238	.184**	-.165*	.190**	.155*	238	.184**	-.165*	.190**	.155*
I20	387	.286**	-.245**	.284**	.254**	303	.318**	-.319**	.319**	.283**
I21	300	.263**	-.240**	.277**	.238**	300	.263**	-.240**	.277**	.238**
I22	207	.163*	-.165*	.162*	.133	207	.163*	-.165*	.162*	.133
I23	206	.170*	-.155*	.176*	.165*	206	.170*	-.155*	.176*	.165*
I24	194	.253**	-.216**	.264**	.245**	194	.253**	-.216**	.264**	.245**
I25	188	.195**	-.181*	.203**	.188**	188	.195**	-.181*	.203**	.188**
I40	380	.132**	-.091	.136**	.097	298	.174**	-.153**	.181**	.143*
I41	282	.230**	-.233**	.239**	.229**	200	.221**	-.217**	.234**	.221**
I42	233	.226**	-.176**	.217**	.214**	183	.235**	-.193**	.228**	.235**
I43	227	.129	-.077	.119	.107	178	.131	-.099	.126	.107
F40	225	.254**	-.251**	.262**	.212**	225	.254**	-.251**	.262**	.212**
F42	209	.243**	-.235**	.246**	.237**	209	.243**	-.235**	.246**	.237**
N40	183	.064	-.030	.070	.072	183	.064	-.030	.070	.072
N41	181	-.015	.002	-.010	-.003	181	-.015	.002	-.010	-.003

Contact_Simulation	399	.057	-.029	.065	.035	310	.078	-.077	.087	.060
Contact_AIRCRAFT	378	.166**	-.134**	.171**	.143**	292	.221**	-.211**	.227**	.196**
Contact_ALL	399	.134**	-.109*	.146**	.119*	310	.167**	-.169**	.180**	.152**
Instruments_Simulation	387	.246**	-.201**	.250**	.223**	303	.268**	-.263**	.278**	.244**
Instruments_AIRCRAFT	380	.171**	-.127*	.174**	.137**	298	.210**	-.195**	.218**	.184**
Instruments_ALL	387	.237**	-.190**	.242**	.206**	303	.272**	-.264**	.283**	.247**
Instruments_BASIC	387	.262**	-.207**	.265**	.225**	303	.303**	-.288**	.312**	.269**
Instruments_RADIO	289	.211**	-.219**	.216**	.204**	207	.198**	-.198**	.207**	.186**
Instruments_NAVIGATION	244	.244**	-.195**	.242**	.227**	194	.261**	-.231**	.266**	.251**
Navigation_AIRCRAFT	183	.037	-.017	.041	.046	183	.037	-.017	.041	.046
Formation_AIRCRAFT	225	.255**	-.252**	.262**	.223**	225	.255**	-.252**	.262**	.223**
Navy Standard Score (NSS)	399	.214**	-.181**	.218**	.183**	310	.254**	-.258**	.263**	.226**

In summary, the ATT appears to measure a largely unidimensional tracking ability that can be studied with both subtest and item level indicators using CTT and IRT methods. This tracking ability correlates meaningfully with various aspects of training performance, as shown by the correlations exceeding .25 in magnitude for several criteria. In fact, the ATT seems to be even more predictive of training performance than was the VTT, which was already identified as a strong candidate for a pilot training selection battery.

AIRPLANE/VERTICAL TRACKING TEST (ATTVTT)
SCORING STRATEGIES AND VALIDITIES

To form "item level" data, we sampled 9 time periods for each task and airplane speed, but increased the duration of each to 7.2 seconds from the 5.2 seconds used previously. As before, we ignored data for one 400ms interval between each time period to reduce score dependencies between adjacent periods. The highest possible score for each time interval was thus 18 for an examinee who was on-target every time a measurement was taken.

Because the VTT and ATT components of this test vary in difficulty, different thresholds were used to transform the continuous data into 5-option polytomous responses for IRT analyses. For the VTT component of the VTTATT, the following categorization scheme was used: 0-1 = 0; 2-3 = 1; 4-5 = 2; 6-7 = 3, 8-18 = 4. For the ATT component, because very few examinees had on-target values larger than 6, a different scheme was used: 0 = 0; 1 = 1; 2 = 2; 3-4 = 3, 5-18 = 4. As before, the main goal of converting continuous ATTVTT data into categorical data is that polytomous IRT models could be applied, making it possible to conduct differential item and test functioning analyses in the future.

Item-Level CTT and IRT Analyses and Results for the ATTVTT

Because the response data were scored polytomously with category codes of higher magnitude indicating better performance on the ATT and VTT components, SGRM (Samejima, 1969) for ordered polytomous responses was chosen for IRT analyses. To verify that the response data for each component of the ATTVTT were sufficiently unidimensional, we conducted separate PCA analyses of the respective inter-item correlations. The scree plot for the ATT analysis is shown in Figure 9, and the scree plot for the VTT analysis is shown in Figure 10. In both cases, the data exhibited a strong first factor. The ratio of first to second eigenvalues exceeded 3.0 for the ATT items, as recommended for application of a unidimensional IRT model (Drasgow & Parsons, 1983; Lord, 1980). The ratio for the VTT items fell slightly short of 3.0, but the elbow in the scree plot is quite pronounced, indicating a strong first factor.

Figure 9. Scree Plot for the Principal Component Analysis of the 9 ATT Items of the ATTVTT

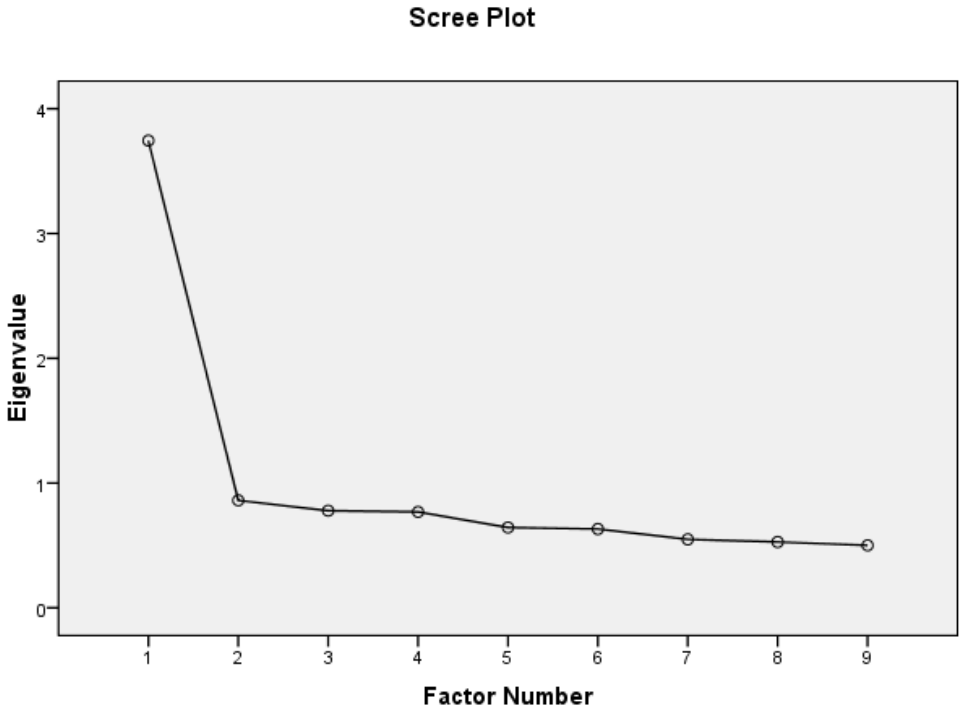
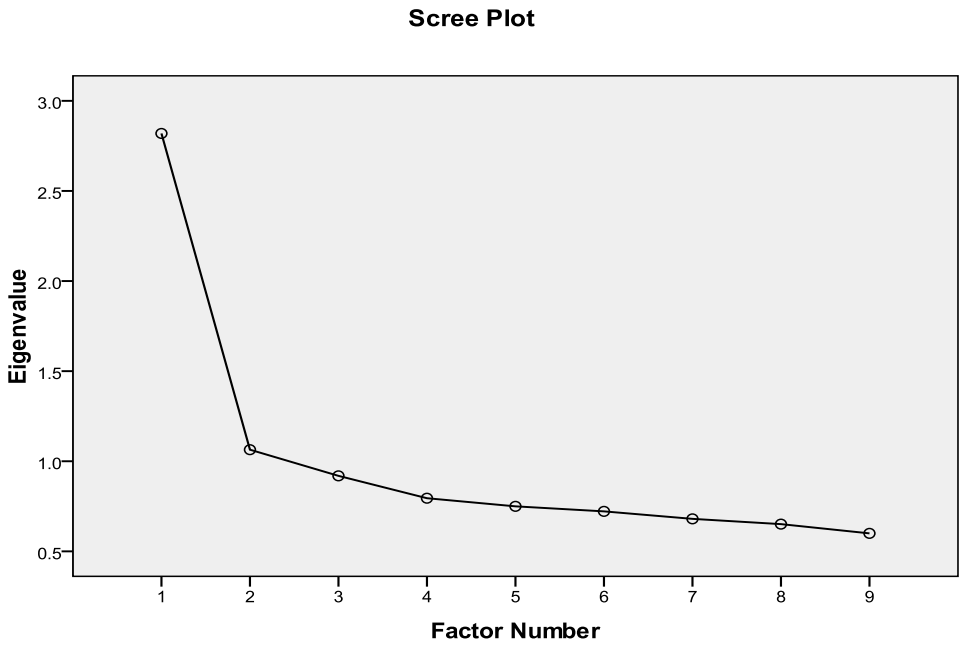


Figure 10. Scree Plot for the Principal Component Analysis of the 9 VTT Items of the ATTVTT



IRT Calibrations of the 9 ATT and 9 VTT Items of the ATTVTT

SGRM item parameters for the ATT and VTT components of ATTVTT were estimated separately using the MULTILOG (Thissen, 1991) computer program (The command files were similar to those shown in previous sections of this report and are therefore omitted). Because the data for each component were coded such that the responses to items fell within one of five ordered categories, there were five SGRM parameters to estimate per item: one discrimination parameter, *a*, and four extremity parameters, b_1, b_2, b_3, and b_4. Scoring and model-data fit analyses were performed using the MODFIT-Z 2.0 computer program (Stark, 2007). Separate parameter estimates, model-data fit statistics, and information functions are reported for the ATT and VTT components of ATTVTT in the tables that follow.

Overall the fit plots and chi-square statistics indicated that SGRM fit the data for both the ATT and VTT components of the ATTVTT very well. As shown in Tables 29 and 30 below, the chi-square statistics were well below the threshold of 3, indicating good fit.

Table 29. Chi-Square Model-Data Fit Statistics for Items Created from the ATT Component Data of the ATTVTT

FREQUENCY TABLE OF ADJUSTED (N=3000) CHISQUARE/DF RATIOS

	<1	1<2	2<3	3<4	4<5	5<7	>7	Mean	SD
Singlets	9	0	0	0	0	0	0	0.00	0.00
Doublets	32	2	2	0	0	0	0	0.21	0.58
Triplets	64	6	7	3	3	1	0	0.73	1.31

Table 30. Chi-Square Model-Data Fit Statistics for Items Created from the VTT Component Data of the ATTVTT

FREQUENCY TABLE OF ADJUSTED (N=3000) CHISQUARE/DF RATIOS

	<1	1<2	2<3	3<4	4<5	5<7	>7	Mean	SD
Singlets	9	0	0	0	0	0	0	0	0.00
Doublets	34	0	2	0	0	0	0	0.23	0.55
Triplets	67	7	5	3	1	1	0	0.63	1.16

Tables 31 and 32 present CTT statistics and IRT parameter estimates for the 9 ATT items and the 9 VTT items of the ATTVTT. Shown are the item means, standard deviations (SD), corrected item-total correlations (CITC), and SGRM item discrimination (a) and extremity parameters (b_1, b_2, b_3, and b_4).

Note that all of the corrected item-total correlations for the ATT component are fairly large, with most being greater than .5, and the IRT *a* parameter estimates for the slow part of the test are in the .85 to.95 range. (These values do not include the 1.7 scaling factor.) The *b4* parameter estimates are also noticeably higher for the medium and fast parts of the test reflecting the increases in difficulty associated with the higher speeds of the target. This is desirable from a

measurement perspective because it means more information is captured by the items at the high end of the trait continuum, leading to better discrimination among high ability examinees.

Table 31. CTT and IRT Statistics for the 9 ATT Items of the ATTVTT

ATT Item Name	Polytomous Responses			SGRM Parameters				
	Mean	SD	CITC	*a*	b_1	b_2	b_3	b_4
ATTVTT_ATT_slow1p	1.73	1.31	.54	0.85	-1.20	-0.06	.69	1.95
ATTVTT_ATT_slow2p	2.02	1.37	.57	0.94	-1.34	-0.45	.30	1.35
ATTVTT_ATT_slow3p	1.98	1.36	.55	0.84	-1.46	-0.37	.35	1.47
ATTVTT_ATT_med1p	1.74	1.30	.55	0.85	-1.20	-0.19	.62	2.14
ATTVTT_ATT_med2p	1.71	1.28	.52	0.76	-1.27	-0.14	.74	2.22
ATTVTT_ATT_med3p	1.76	1.25	.51	0.73	-1.53	-0.19	.71	2.37
ATTVTT_ATT_fast1p	1.46	1.23	.52	0.74	-1.04	0.26	1.11	2.81
ATTVTT_ATT_fast2p	1.50	1.20	.47	0.65	-1.20	0.18	1.18	3.23
ATTVTT_ATT_fast3p	1.61	1.24	.52	0.77	-1.19	-0.05	.93	2.57

In Table 32, it can be seen that the corrected item-total correlations for the VTT component of the ATTVTT are good, but clearly lower than for the ATT component. In this case, most of the correlations were in the .35 to .45 range and the *a* parameters were relatively small. (The *a* parameters do not include the 1.7 scaling factor.). Interestingly, although the VTT is arguably an easier task than the ATT when performed individually, several b_4 parameters in Table 32 are above 3.5 at medium and fast speeds, suggesting that VTT became harder for examinees when performed simultaneously with ATT in this subtest. One possibility that requires further investigation is whether examinees become so consumed by efforts to perform well on the ATT component that the VTT component is essentially left unattended due to cognitive overload. It would be interesting to investigate whether such effects diminish with additional experience in a flight simulator or flight hours, as the target visualizations and independent manipulations of the throttle and joystick presumably become more automated.

Table 32. CTT and IRT Statistics for the 9 VTT Items of the ATTVTT

VTT Item Name	Polytomous Responses			SGRM Parameters				
	Mean	SD	CITC	*a*	b_1	b_2	b_3	b_4
ATTVTT_VTT_slow1p	1.83	1.18	.34	0.47	-2.91	-0.27	1.40	2.72
ATTVTT_VTT_slow2p	1.90	1.24	.44	0.64	-2.24	-0.31	.88	1.93
ATTVTT_VTT_slow3p	1.90	1.22	.44	0.69	-2.02	-0.36	.80	2.02
ATTVTT_VTT_med1p	1.41	1.04	.38	0.55	-1.82	0.47	2.08	3.89
ATTVTT_VTT_med2p	1.40	1.05	.40	0.62	-1.55	0.37	2.02	3.42
ATTVTT_VTT_med3p	1.49	1.09	.43	0.70	-1.66	0.35	1.56	2.84
ATTVTT_VTT_fast1p	1.26	1.03	.33	0.49	-1.51	0.88	2.46	4.95
ATTVTT_VTT_fast2p	1.32	1.05	.38	0.56	-1.47	0.49	2.35	3.87
ATTVTT_VTT_fast3p	1.27	.97	.43	0.68	-1.35	0.65	2.15	3.93

Figure 11. Test Information Function for the 9 ATT Items of the ATTVTT

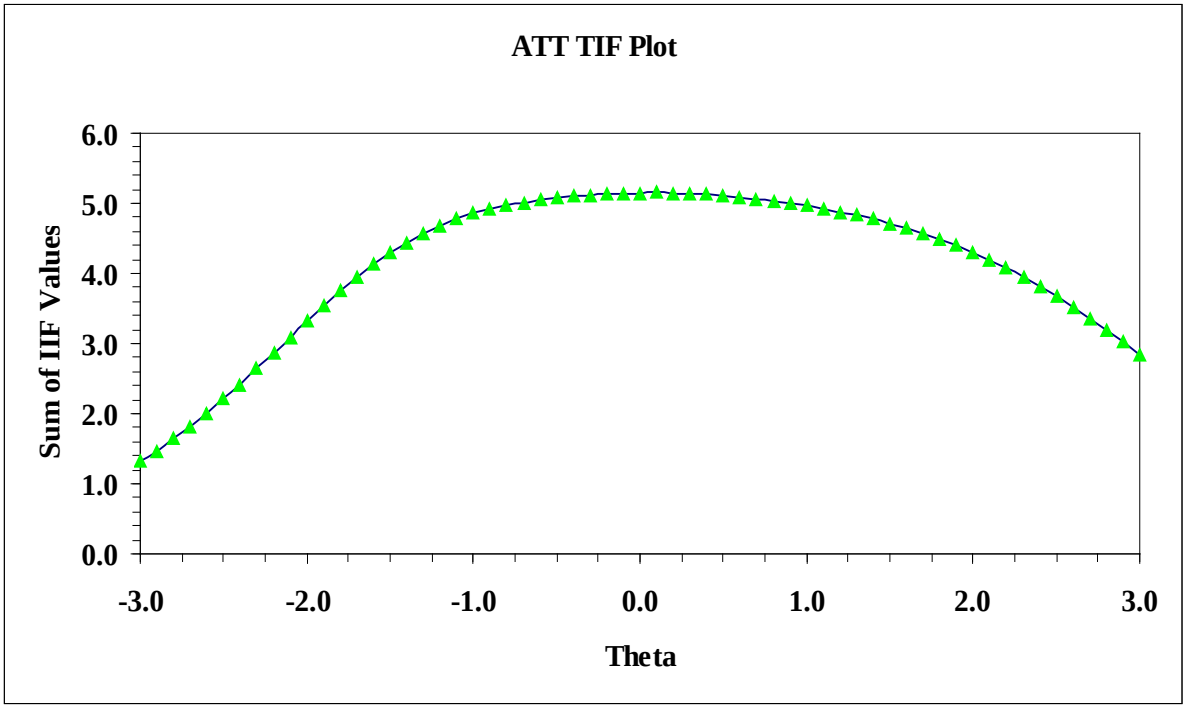
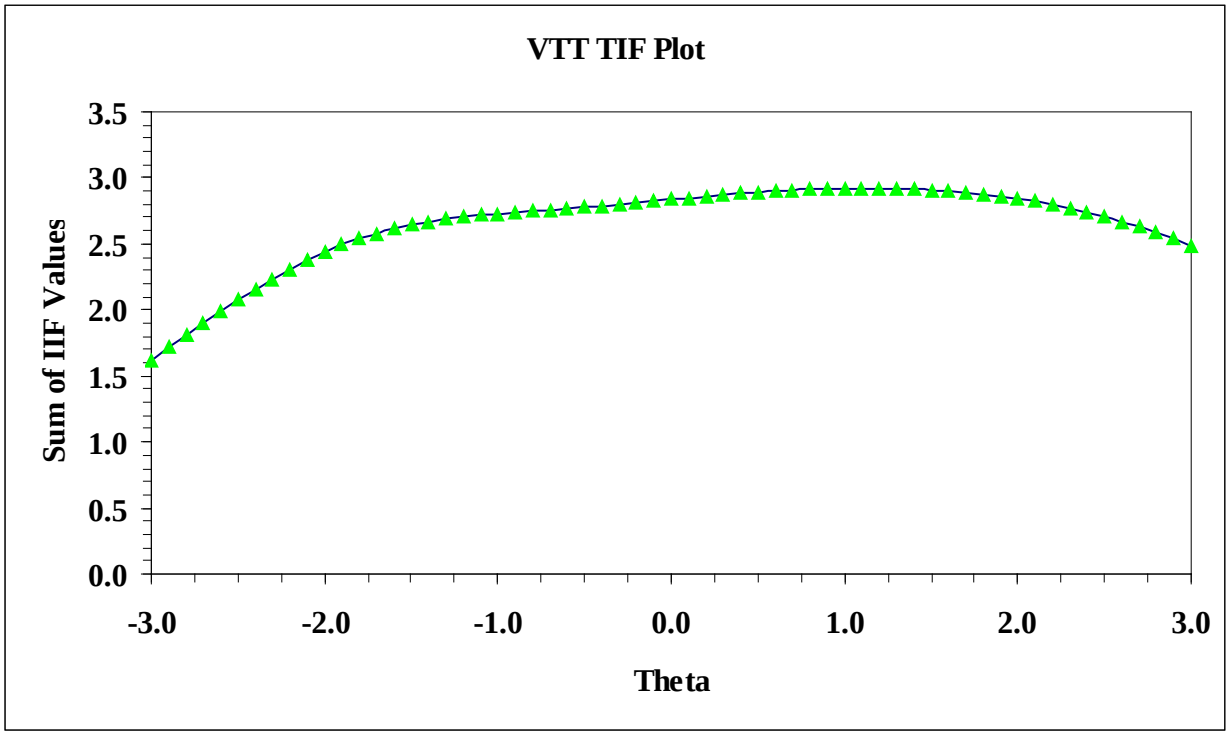


Figure 12. Test Information Function for the 9 VTT Items of the ATTVTT



ATTVTT Scale Scores

The subtest level indicators for the ATT and VTT components of the ATTVTT were also examined separately for SPs, SNFOs, and the total sample. The total number of redirects (ATTVTT ATT Redirects and ATTVTT VTT Redirects), the average distance between the respective crosshairs and the targets during the test (ATTVTT ATT Average Distance and ATTVTT VTT Average Distance), the total numbers of on-target responses (ATTVTT ATT Total On Target and ATTVTT VTT Total On Target), and the IRT scores (ATTVTT ATT IRT Score and ATTVTT VTT IRT Score) are all indicators of examinee ability. As with the individually administered ATT and VTT assessments, the average distance measures are negatively related to the scores for the other components in the ATTVTT. The ATT and VTT components themselves correlated .45 to .55 (see Table 35).

Tables 33 and 34 show descriptive statistics for the ATT and VTT components of the ATTVTT when analyzed in the total sample and across SP and SNFO student groups. As with the previous tests in the PBM sequence, SPs performed better on the ATT component by about a third of the total sample SD and the effect size was somewhat smaller for the VTT component.

Table 33. Performance on the ATT Component of the ATTVTT Across SP and SNFO Student Groups

		ATTVTT ATT Redirects		ATTVTT ATT Average Distance		ATTVTT ATT Total On Target		ATTVTT ATT IRT Score	
Program	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SNFO	89	7.73	4.92	125.85	36.69	28.67	17.16	-0.26	0.91
SP	310	9.83	5.57	111.53	34.74	36.21	19.39	0.07	0.89
Total	399	9.36	5.50	114.72	35.64	34.53	19.15	0.00	0.90

Table 34. Performance on the VTT Component of the ATTVTT Across SP and SNFO Student Groups

		ATTVTT VTT Redirects		ATTVTT VTT Average Distance		ATTVTT VTT Total On Target		ATTVTT VTT IRT Score	
Program	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SNFO	89	11.92	4.42	92.30	25.54	57.16	20.03	-0.17	0.84
SP	310	12.69	4.52	86.28	23.62	61.44	20.98	0.04	0.83
Total	399	12.52	4.51	87.62	24.16	60.48	20.82	0.00	0.84

Table 35. Correlations Between the ATT and VTT Component Scores of the ATTVTT

	ATTVTT ATT Redirects	ATTVTT ATT Average Distance	ATTVTT ATT Total On Target	ATTVTT ATT IRT Score	ATTVTT VTT Redirects	ATTVTT VTT Average Distance	ATTVTT VTT Total On Target	ATTVTT VTT IRT Score
ATTVTT ATT Redirects	1	-.868**	.986**	.923**	.565**	-.562**	.559**	.502**
ATTVTT ATT Average Distance	-.868**	1	-.868**	-.869**	-.458**	.528**	-.456**	-.443**
ATTVTT ATT Total On Target	.986**	-.868**	1	.939**	.555**	-.555**	.550**	.492**
ATTVTT ATT IRT Score	.923**	-.869**	.939**	1	.485**	-.508**	.479**	.446**
ATTVTT VTT Redirects	.565**	-.458**	.555**	.485**	1	-.873**	.985**	.879**
ATTVTT VTT Average Distance	-.562**	.528**	-.555**	-.508**	-.873**	1	-.863**	-.834**
ATTVTT VTT Total On Target	.559**	-.456**	.550**	.479**	.985**	-.863**	1	.894**

ATTVTT VTT IRT Score	.502**	-.443**	.492**	.446**	.879**	-.834**	.894**	1

Table 36 shows the correlations of the ATTVTT ATT component scores with other potential predictors of training performance. All of the ATT component scores showed sizeable correlations (.38 to .43 in magnitude) with simExperience and small correlations ($|r| < .1$) with aTraining and Education. Correlations with AQR_ and PFAR_ variables were also noteworthy, with magnitudes exceeding .3.

Table 36. Correlations Between the ATT Scores of the ATTVTT and Other Predictors

	N	Mean	SD	ATTVTT ATT Redirects	ATTVTT ATT Average Distance	ATTVTT ATT Total On Target	ATTVTT ATT IRT Score
aTraining	390	0.23	0.72	.081	-.067	.088	.088
Education	385	2.88	0.59	.102*	-.082	.101*	.082
simExperience	399	0.74	0.83	.432**	-.376**	.428**	.389**
flightHours	391	0.69	1.38	.151**	-.101*	.158**	.132**
ANI_RAW	332	0.58	0.53	.212**	-.247**	.225**	.210**
MST_RAW	332	0.34	0.67	.116*	-.113*	.120*	.111*
RCT_RAW	332	0.43	0.53	.052	-.057	.053	.034
SAT_Post2004	332	0.76	0.64	.239**	-.235**	.244**	.233**
MCT_Post2004	332	0.50	0.64	.219**	-.225**	.216**	.189**
AQR_Post2004	332	0.55	0.52	.288**	-.306**	.294**	.267**
PFAR_Post2004	332	0.67	0.50	.316**	-.339**	.326**	.301**
FOFAR_Post2004	332	0.65	0.53	.276**	-.288**	.286**	.267**
OAR_Post2004	332	0.50	0.62	.212**	-.216**	.212**	.186**

Table 37 shows the correlations of the ATTVTT VTT component scores with other potential predictors of training performance. The VTT component scores showed smaller correlations (.20s) than did the ATT component scores (.30s) with simExperience and virtually no correlation with aTraining and Education. However, correlations with AQR_, PFAR_, and FOFAR_ variables were moderate, with magnitudes in the .25 to .35 range.

Table 37. Correlations Between the VTT Scores of the ATTVTT and Other Predictors

	N	Mean	SD	ATTVTT VTT Redirects	ATTVTT VTT Average Distance	ATTVTT VTT Total On Target	ATTVTT VTT IRT Score
aTraining	390	0.23	0.72	.002	-.023	.008	-.017
Education	385	2.88	0.59	.023	-.029	.033	.035
simExperience	399	0.74	0.83	.246**	-.247**	.232**	.191**
flightHours	391	0.69	1.38	.023	-.019	.021	.003
ANI_RAW	332	0.58	0.53	.157**	-.201**	.152**	.167**

MST_RAW	332	0.34	0.67	.169**	-.198**	.170**	.177**
RCT_RAW	332	0.43	0.53	.089	-.123*	.094	.098
SAT_Post2004	332	0.76	0.64	.188**	-.267**	.196**	.196**
MCT_Post2004	332	0.50	0.64	.164**	-.212**	.164**	.171**
AQR_Post2004	332	0.55	0.52	.250**	-.320**	.250**	.265**
PFAR_Post2004	332	0.67	0.50	.248**	-.326**	.248**	.261**
FOFAR_Post2004	332	0.65	0.53	.267**	-.347**	.271**	.284**
OAR_Post2004	332	0.50	0.62	.195**	-.245**	.196**	.204**

Tables 38 and 39 present the correlations between the ATT and VTT component scores of the ATTVTT and training criteria (block grades and training composites) for the total sample and for student pilots only. As can be seen in Table 38, the ATT component scores showed correlations with criteria I20, I21, and F42 in the .25 to .30 range for SPs, and average correlations with the composite criteria, Contact_ALL and Instruments_ALL, of .17 and .29 respectively. The latter is quite substantial and suggests good utility for decision making.

The correlations in Table 39 for the VTT component scores of the ATTVTT show similar patterns, although the magnitudes are somewhat smaller than those for the ATT component scores. This is not surprising given the smaller discrimination parameters and effect size differences across SPs and SNFOs shown in previous tables. Nonetheless, the correlations with I20, I21, F42, and Instruments_BASIC are good, with values in the mid .20s, indicating that the VTT component scores do provide useful information for predictive purposes.

Table 38. Correlations Between the ATT Component Scores of the ATTVTT and Navy Pilot Training Criteria

	Total Sample					Students Pilots (SPs)				
Training Block Name	N	ATTVTT ATT Redirects	ATTVTT ATT Average Distance	ATTVTT ATT Total On Target	ATTVTT ATT IRT Score	N	ATTVTT ATT Redirects	ATTVTT ATT Average Distance	ATTVTT ATT Total On Target	ATTVTT ATT IRT Score
C20	399	.092	-.077	.096	.060	310	.118*	-.112*	.119*	.084
C40	86	.073	-.013	.076	.054	-	-	-	-	-
C41	374	.136**	-.060	.136**	.126*	292	.181**	-.124*	.182**	.175**
C42	367	.094	-.087	.094	.082	284	.157**	-.128*	.152*	.138*
C43	270	.087	-.098	.082	.073	270	.087	-.098	.082	.073
C45	262	.162**	-.195**	.172**	.169**	262	.162**	-.195**	.172**	.169**
C46	246	.160*	-.179**	.162*	.134*	246	.160*	-.179**	.162*	.134*
C47	238	.156*	-.101	.165*	.125	238	.156*	-.101	.165*	.125
I20	387	.276**	-.248**	.283**	.271**	303	.307**	-.303**	.310**	.300**
I21	300	.257**	-.256**	.266**	.243**	300	.257**	-.256**	.266**	.243**
I22	207	.120	-.179**	.123	.114	207	.120	-.179**	.123	.114
I23	206	.141*	-.172*	.136	.121	206	.141*	-.172*	.136	.121
I24	194	.200**	-.223**	.201**	.169*	194	.200**	-.223**	.201**	.169*
I25	188	.166*	-.183*	.163*	.139	188	.166*	-.183*	.163*	.139
I40	380	.143**	-.154**	.141**	.131*	298	.162**	-.168**	.158**	.143*
I41	282	.226**	-.232**	.227**	.209**	200	.231**	-.231**	.231**	.227**
I42	233	.180**	-.157*	.198**	.156*	183	.207**	-.185*	.222**	.165*
I43	227	.077	-.098	.082	.079	178	.046	-.060	.040	.042
F40	225	.199**	-.221**	.207**	.189**	225	.199**	-.221**	.207**	.189**
F42	209	.272**	-.283**	.267**	.268**	209	.272**	-.283**	.267**	.268**
N40	183	.079	-.014	.090	.023	183	.079	-.014	.090	.023
N41	181	-.024	.042	-.005	-.027	181	-.024	.042	-.005	-.027

Contact_Simulation	399	.092	-.077	.096	.060	310	.118*	-.112*	.119*	.084
Contact_AIRCRAFT	378	.174**	-.127*	.173**	.148**	292	.223**	-.185**	.221**	.197**
Contact_ALL	399	.157**	-.119*	.160**	.127*	310	.192**	-.163**	.192**	.161**
Instruments_Simulation	387	.245**	-.221**	.252**	.235**	303	.272**	-.274**	.275**	.255**
Instruments_AIRCRAFT	380	.196**	-.192**	.195**	.173**	298	.230**	-.223**	.225**	.202**
Instruments_ALL	387	.250**	-.240**	.254**	.234**	303	.282**	-.283**	.283**	.262**
Instruments_BASIC	387	.259**	-.247**	.265**	.251**	303	.290**	-.293**	.294**	.278**
Instruments_RADIO	289	.185**	-.219**	.185**	.166**	207	.178*	-.219**	.176*	.171*
Instruments_NAVIGATION	244	.198**	-.220**	.207**	.176**	194	.211**	-.238**	.212**	.174*
Navigation_AIRCRAFT	183	.027	.016	.042	-.006	183	.027	.016	.042	-.006
Formation_AIRCRAFT	225	.233**	-.262**	.238**	.232**	225	.233**	-.262**	.238**	.232**
Navy Standard Score (NSS)	399	.233**	-.220**	.240**	.218**	310	.267**	-.262**	.270**	.247**

Table 39. Correlations Between the VTT Component Scores of the ATTVTT and Navy Pilot Training Criteria

	Total Sample					Students Pilots (SPs)				
Training Block Name	N	ATTVTT VTT Redirects	ATTVTT VTT Average Distance	ATTVTT VTT Total On Target	ATTVTT VTT IRT Score	N	ATTVTT VTT Redirects	ATTVTT VTT Average Distance	ATTVTT VTT Total On Target	ATTVTT VTT IRT Score
C20	399	.030	-.069	.033	.061	310	.038	-.067	.046	.055
C40	86	.096	-.164	.087	.107	-	-	-	-	-
C41	374	.050	-.080	.045	.021	292	.079	-.104	.078	.055
C42	367	.032	-.089	.029	.042	284	.118*	-.171**	.114	.109
C43	270	.083	-.084	.076	.056	270	.083	-.084	.076	.056
C45	262	.104	-.101	.087	.098	262	.104	-.101	.087	.098
C46	246	.064	-.083	.038	.046	246	.064	-.083	.038	.046
C47	238	.136*	-.116	.124	.083	238	.136*	-.116	.124	.083

I20	387	.202**	-.235**	.194**	.167**	303	.209**	-.252**	.201**	.179**
I21	300	.190**	-.228**	.182**	.157**	300	.190**	-.228**	.182**	.157**
I22	207	.055	-.100	.052	.011	207	.055	-.100	.052	.011
I23	206	.049	-.094	.040	.002	206	.049	-.094	.040	.002
I24	194	.162*	-.168*	.154*	.100	194	.162*	-.168*	.154*	.100
I25	188	.055	-.128	.061	.030	188	.055	-.128	.061	.030
I40	380	.191**	-.181**	.186**	.187**	298	.210**	-.188**	.206**	.201**
I41	282	.091	-.157**	.092	.090	200	.108	-.109	.105	.084
I42	233	.129*	-.135*	.125	.127	183	.135	-.142	.133	.143
I43	227	.083	-.105	.064	.054	178	.087	-.088	.063	.059
F40	225	.174**	-.204**	.188**	.179**	225	.174**	-.204**	.188**	.179**
F42	209	.206**	-.227**	.209**	.198**	209	.206**	-.227**	.209**	.198**
N40	183	.100	-.114	.104	.106	183	.100	-.114	.104	.106
N41	181	.081	-.052	.081	.056	181	.081	-.052	.081	.056
Contact_Simulation	399	.030	-.069	.033	.061	310	.038	-.067	.046	.055
Contact_AIRCRAFT	378	.087	-.142**	.078	.083	292	.113	-.158**	.104	.099
Contact_ALL	399	.063	-.112*	.057	.074	310	.070	-.108	.068	.073
Instruments_Simulation	387	.169**	-.211**	.165**	.119*	303	.167**	-.220**	.164**	.119*
Instruments_AIRCRAFT	380	.168**	-.170**	.165**	.154**	298	.191**	-.178**	.189**	.172**
Instruments_ALL	387	.180**	-.215**	.176**	.140**	303	.190**	-.225**	.187**	.146*
Instruments_BASIC	387	.225**	-.247**	.217**	.190**	303	.240**	-.264**	.232**	.205**
Instruments_RADIO	289	.067	-.157**	.066	.055	207	.076	-.111	.071	.037
Instruments_NAVIGATION	244	.122	-.158*	.113	.076	194	.128	-.159*	.118	.079
Navigation_AIRCRAFT	183	.105	-.104	.104	.092	183	.105	-.104	.104	.092
Formation_AIRCRAFT	225	.183**	-.219**	.195**	.192**	225	.183**	-.219**	.195**	.192**
Navy Standard Score (NSS)	399	.154**	-.202**	.151**	.131**	310	.171**	-.211**	.171**	.140*

Table 40 presents the results of multiple regression analyses using the ATT and VTT components of the ATTVTT as predictors of composite performance criteria for 310 Student Pilots. We focused on SPs because performance on complex tracking tasks is less likely to be relevant for SNFOs. Consistent with expectations based on the zero-order correlations, good predictive validities were found for the Instruments, Formation, and NSS criteria with multiple R values in the 0.25 to 0.30 range. Predictive validities were smaller though for Contact and Navigation criteria with multiple R values at or below 0.20.

In nearly all cases, the ATT component of the ATTVTT had stronger and significant relations with criteria. This is not particularly surprising given the somewhat lower discrimination parameters observed for the VTT items in the IRT analyses. Perhaps it is too difficult to maintain an adequate on target hit rate on the VTT task while attending to the ATT task.

We also conducted moderated regression analyses by adding an interaction term for the standardized ATT and VTT components. Nineteen of twenty of these moderated regression analyses showed no significant interaction between ATT and VTT. In the one case involving ATT and VTT IRT scores predicting Instruments_All grades, the interaction term was significant although the effect size was fairly small (change in R^2 was just 0.013). As can be seen from Figure 13, which depicts this interaction, examinees with superior performance on the ATT component tend to have better Instruments grades regardless of their VTT performance; and examinees performing both tasks well tend to have the highest grades. Overall, because only one statistically significant interaction was found and twenty significance tests were conducted, this result could be due to chance and should not be over-emphasized.

Figure 13. Interaction Between the Standardized ATT IRT Scores and the Standardized VTT IRT Scores When Predicting the Instruments_All Training Composite for SPs

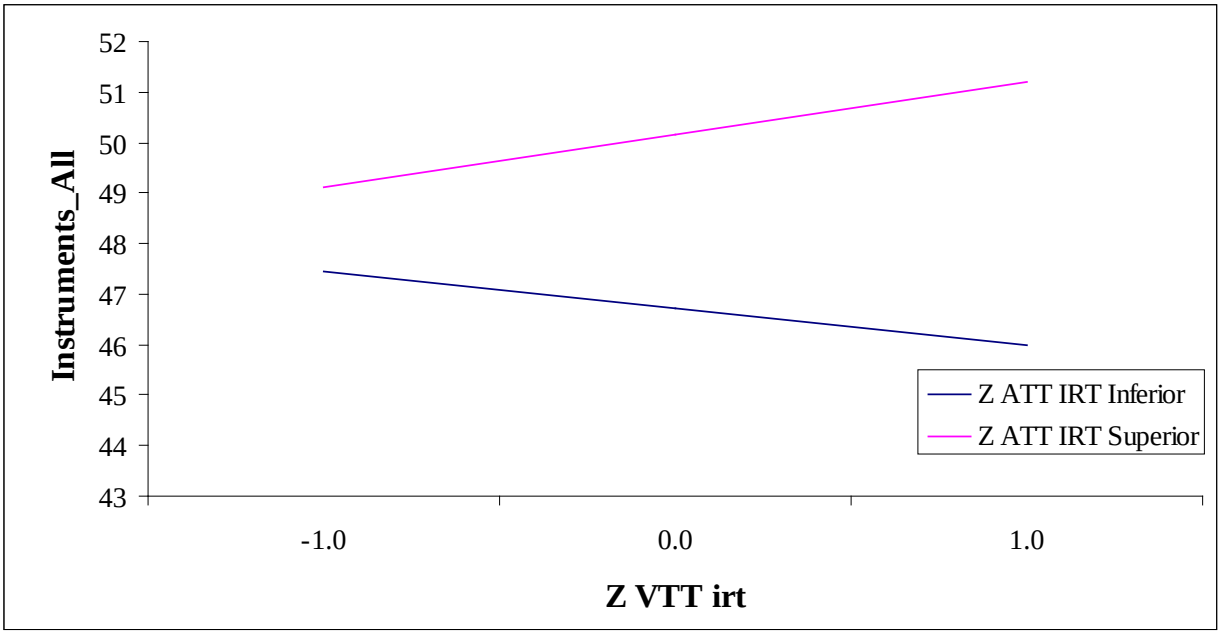


Table 40. ATTVTT Multiple Regression Results using the ATT and VTT Component Scores as Predictors of Navy Pilot Training Criteria

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	R
		B	Std. Error	Beta			
Contact ALL							
	(Constant)	47.50	1.22		38.80	0.00	0.198
	AttVtt VTT Redirects	-0.09	0.11	-0.06	-0.86	0.39	
	AttVtt ATT Redirects	0.30	0.09	0.23	3.31	0.00	
	(Constant)	53.54	1.69		31.67	0.00	0.166
	AttVtt VTT Average Distance	-0.01	0.02	-0.03	-0.51	0.61	
	AttVtt ATT Average Distance	-0.03	0.01	-0.15	-2.23	0.03	
	(Constant)	47.33	1.28		37.10	0.00	0.198
	ATTVTT VTT Total On Target	-0.02	0.02	-0.05	-0.81	0.42	
	ATTVTT ATT Total On Target	0.08	0.03	0.22	3.32	0.00	
	(Constant)	49.11	0.41		118.73	0.00	0.161
	ATTVTT VTT IRT Score	0.03	0.55	0.00	0.06	0.95	
	ATTVTT ATT IRT Score	1.32	0.52	0.16	2.55	0.01	
Instruments ALL							
	(Constant)	44.19	1.36		32.39	0.00	0.284
	AttVtt VTT Redirects	0.08	0.12	0.05	0.68	0.50	
	AttVtt ATT Redirects	0.38	0.10	0.26	3.82	0.00	
	(Constant)	58.26	1.86		31.28	0.00	0.298

	AttVtt VTT Average Distance	-0.04	0.02	-0.11	-1.72	0.09	
	AttVtt ATT Average Distance	-0.05	0.02	-0.23	-3.56	0.00	
	(Constant)	43.86	1.42		30.88	0.00	0.286
	ATTVTT VTT Total On Target	0.02	0.03	0.05	0.73	0.47	
	ATTVTT ATT Total On Target	0.11	0.03	0.26	3.90	0.00	
	(Constant)	48.79	0.46		106.34	0.00	0.265
	ATTVTT VTT IRT Score	0.41	0.61	0.04	0.67	0.50	
	ATTVTT ATT IRT Score	2.25	0.57	0.24	3.96	0.00	
Navigation AIRCRAFT							
	(Constant)	47.84	1.98		24.16	0.00	0.109
	AttVtt VTT Redirects	0.24	0.17	0.12	1.42	0.16	
	AttVtt ATT Redirects	-0.05	0.13	-0.03	-0.39	0.70	
	(Constant)	52.65	2.60		20.25	0.00	0.128
	AttVtt VTT Average Distance	-0.05	0.03	-0.14	-1.72	0.09	
	AttVtt ATT Average Distance	0.02	0.02	0.09	1.02	0.31	
	(Constant)	47.67	2.09		22.83	0.00	0.104
	ATTVTT VTT Total On Target	0.05	0.04	0.11	1.28	0.20	
	ATTVTT ATT Total On Target	0.00	0.04	-0.01	-0.10	0.92	
	(Constant)	50.32	0.64		79.04	0.00	0.102
	ATTVTT VTT IRT Score	1.12	0.82	0.11	1.37	0.17	

	ATTVTT ATT IRT Score	-0.46	0.76	-0.05	-0.60	0.55	
Formation AIRCRAFT	(Constant)	44.87	1.85		24.22	0.00	0.244
	AttVtt VTT Redirects	0.18	0.16	0.09	1.14	0.25	
	AttVtt ATT Redirects	0.30	0.12	0.19	2.49	0.01	
	(Constant)	59.66	2.35		25.34	0.00	0.280
	AttVtt VTT Average Distance	-0.05	0.03	-0.12	-1.56	0.12	
	AttVtt ATT Average Distance	-0.05	0.02	-0.20	-2.71	0.01	
	(Constant)	44.09	1.93		22.88	0.00	0.255
	ATTVTT VTT Total On Target	0.05	0.03	0.11	1.42	0.16	
	ATTVTT ATT Total On Target	0.09	0.03	0.19	2.52	0.01	
	(Constant)	49.94	0.58		86.57	0.00	0.255
	ATTVTT VTT IRT Score	1.25	0.77	0.12	1.62	0.11	
	ATTVTT ATT IRT Score	1.79	0.69	0.18	2.58	0.01	
Navy Standard Score (NSS)	(Constant)	44.18	1.59		27.76	0.00	0.268
	AttVtt VTT Redirects	0.06	0.14	0.03	0.42	0.68	
	AttVtt ATT Redirects	0.44	0.12	0.25	3.75	0.00	
	(Constant)	59.42	2.18		27.28	0.00	0.277
	AttVtt VTT Average Distance	-0.04	0.03	-0.11	-1.65	0.10	
	AttVtt ATT Average Distance	-0.06	0.02	-0.21	-3.26	0.00	

(Constant)	43.75	1.66		26.41	0.00	0.271
ATTVTT VTT Total On Target	0.01	0.03	0.03	0.49	0.62	
ATTVTT ATT Total On Target	0.13	0.03	0.25	3.84	0.00	
(Constant)	49.03	0.54		91.35	0.00	0.249
ATTVTT VTT IRT Score	0.47	0.71	0.04	0.65	0.51	
ATTVTT ATT IRT Score	2.51	0.67	0.23	3.74	0.00	

In summary, the ATTVTT appears to be a challenging test for examinees and it predicts several performance criteria well. When the ATT and VTT components are analyzed separately, each shows a single dominant dimension underlying the item responses and the polytomized distance data can be fit quite well using Samejima's (1969) graded response IRT model. Interestingly, the SGRM analyses showed clear increases in the difficulty of ATT items as the speed of the target increased, but the pattern was not evident for the VTT items, perhaps because they were less discriminating. Whether or not the lower discriminations were due to increased attention to the ATT component of this subtest is something that deserves further study.

MULTITRACKING TEST (MTT)
SCORING STRATEGIES AND VALIDITIES

In the MTT (also referred as AttVttDlt), examinees must perform dichotic listening, one-dimensional tracking, and two-dimensional tracking tasks simultaneously. Clearly this places high cognitive demands on a respondent.

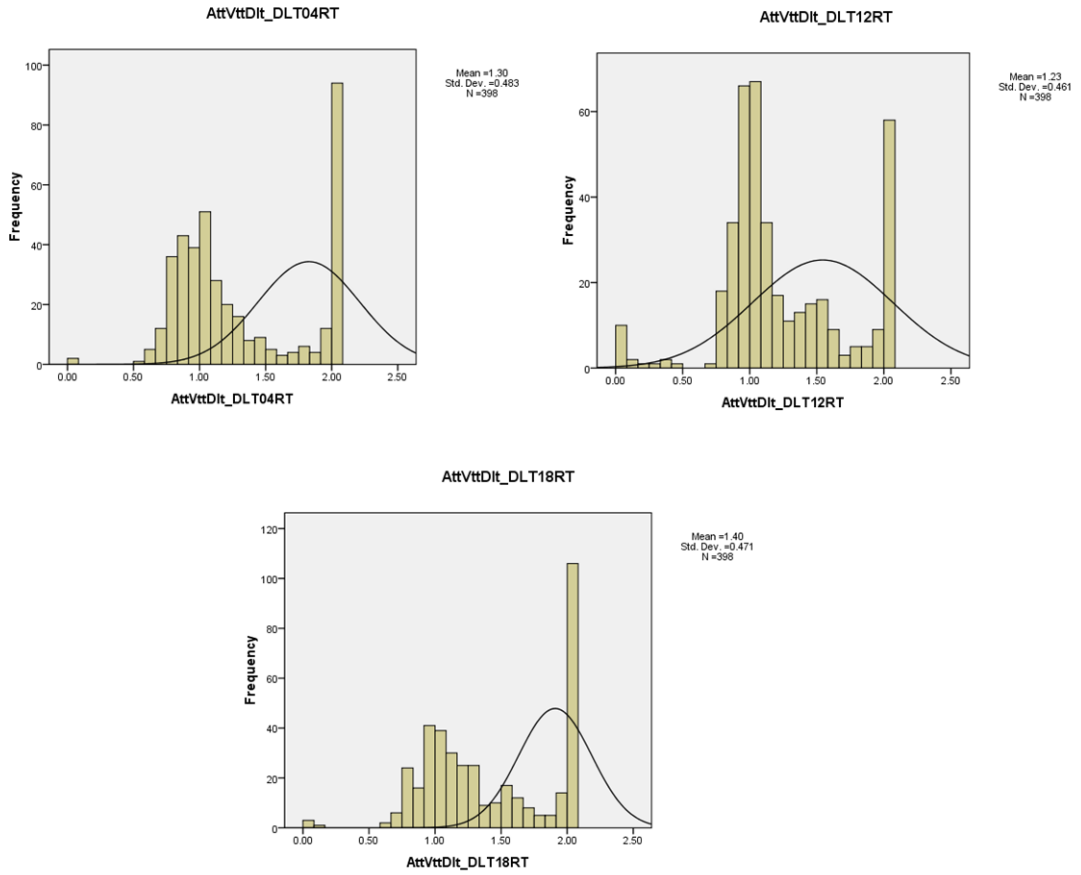
The mechanics of the MTT are the same as in the ATTVTT, except that the examinee must also manipulate the trigger on a joystick and the RDR Cursor button on a throttle in response to odd or even numbers presented through headphones in a "target ear". The same types of performance data are recorded in MTT as when the DLT and ATTVTT are administered separately, with one important difference being that the MTT is 180 seconds in duration.

To form "item level" data for IRT analyses of the ATT and VTT subcomponents, we sampled 9 time periods for each task and airplane speed, but used the 7.2 second (18 400ms interval) time period for derivation of each polytomous item. As before, we ignored data for one 400ms interval between each period to reduce score dependencies between adjacent intervals. The highest possible score for each time interval was thus 18 for an examinee who was on-target every time a measurement was taken.

Different thresholds were used to transform the continuous tracking data into 5-option polytomous responses for the IRT analyses. For the VTT component of the MTT, the following categorization scheme was used: 0-1 = 0; 2-3 = 1; 4-5 = 2; 6-7 = 3, 8-18 = 4. For the ATT component, because very few examinees had on-target values larger than 6, a different scheme was used: 0 = 0; 1 = 1; 2 = 2; 3-4 = 3, 5-18 = 4.

For reasons discussed above in the DLT section, the dichotic listening data recorded during MTT could not be analyzed using IRT methods. Response time distributions for three illustrative items are presented in Figure 14. The histograms show that the response times were positively skewed, with an unusually high peak for the 2000ms category. These values represent omitted responses and response latencies of exactly 2000ms. There is no way to differentiate between the two in this dataset. As Figure 14 indicates, latencies near 2000ms in the DLT component of the MTT appeared to be more prevalent than in the DLT administered in isolation, which is not surprising.

Figure 14. Response Time Distributions for Three Illustrative Dichotic Listening Items in the MTT



Item-Level CTT and IRT Analyses and Results for the ATT and VTT Components of the MTT

Because the response data were scored polytomously with category codes of higher magnitude indicating better performance on the ATT and VTT components, SGRM (Samejima, 1969) for ordered polytomous responses was chosen for IRT analyses. To verify that the response data for each component was sufficiently unidimensional, we conducted separate principal component analyses of the ATT and VTT inter-item correlations. The scree plot for the ATT analysis is shown in Figure 15, and the scree plot for the VTT analysis is shown in Figure 16. In both cases, the data exhibited a strong first factor with the ratio of first to second eigenvalues exceeding 3.0 as recommended for application of a unidimensional IRT model (Drasgow & Parsons, 1983; Lord, 1980).

Figure 15. Scree Plot for the Principal Component Analysis of the 9 ATT Items of the MTT

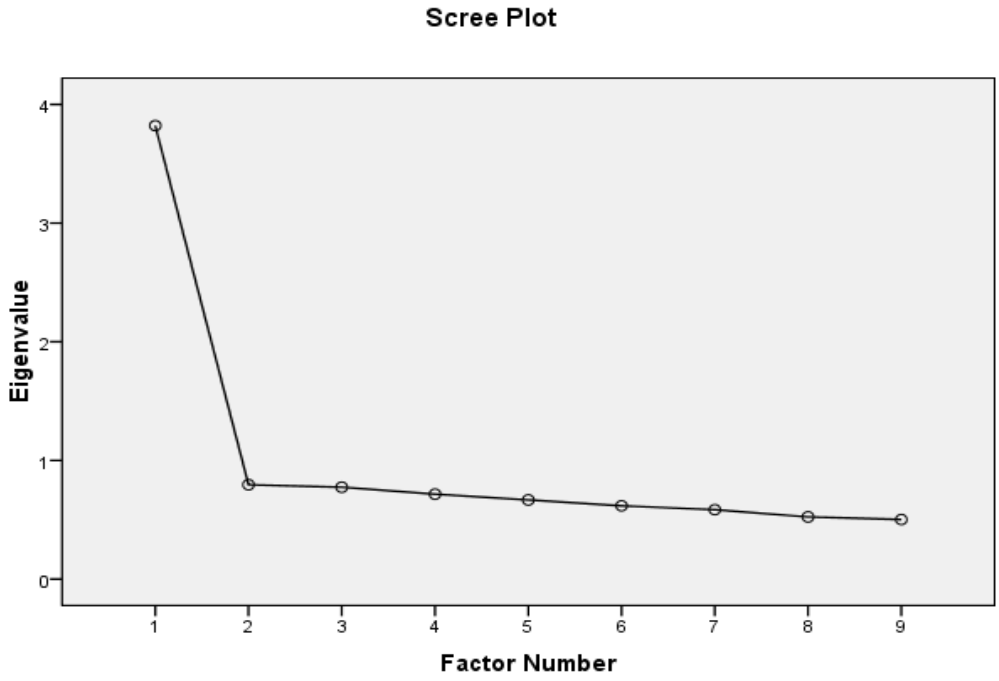
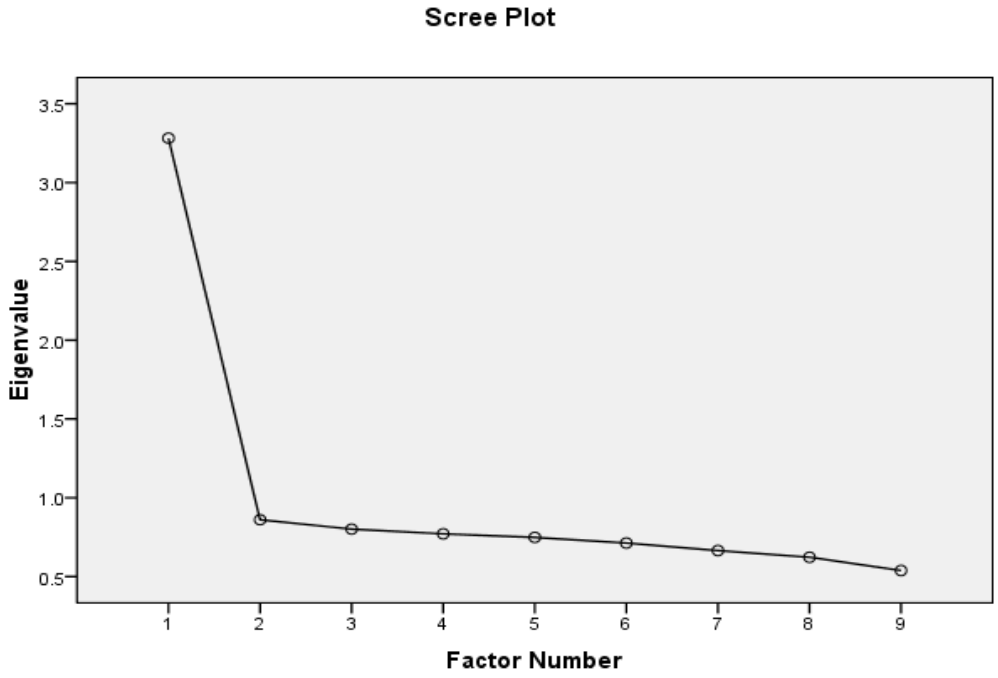


Figure 16. Scree Plot for the Principal Component Analysis of the 9 VTT Items of the MTT



IRT Calibrations of the 9 ATT and 9 VTT Items of the MTT

SGRM item parameters for the ATT and VTT components of the MTT were estimated separately using the MULTILOG (Thissen, 1991) computer program (The command files were similar to those shown in previous sections of this report.). Because the data for each component were coded such that the responses to items fell within one of five ordered categories, there were five SGRM parameters to estimate per item: one discrimination parameter, *a*, and four extremity parameters, b_1, b_2, b_3, and b_4. Scoring and model-data fit analyses were performed using the MODFIT-Z 2.0 computer program (Stark, 2007). Separate parameter estimates, model-data fit statistics, and information functions are reported for the ATT and VTT components of the MTT in the tables that follow.

Overall the fit plots and chi-square statistics indicated that SGRM fit the data for both the ATT and VTT components of the MTT very well. As shown in Tables 41 and 42, the chi-square statistics were well below the threshold of 3, indicating good fit.

Table 41. Chi-Square Model-Data Fit Statistics for Items Created from the ATT Component Data of the MTT

FREQUENCY TABLE OF ADJUSTED (N=3000) CHISQUARE/DF RATIOS

	<1	1<2	2<3	3<4	4<5	5<7	>7	Mean	SD
Singlets	9	0	0	0	0	0	0	0	0
Doublets	36	0	0	0	0	0	0	0.02	0.14
Triplets	75	8	0	1	0	0	0	0.27	0.56

Table 42. Chi-Square Model-Data Fit Statistics for Items Created from the VTT Component Data of the MTT

FREQUENCY TABLE OF ADJUSTED (N=3000) CHISQUARE/DF RATIOS

	<1	1<2	2<3	3<4	4<5	5<7	>7	Mean	SD
Singlets	9	0	0	0	0	0	0	0	0
Doublets	35	1	0	0	0	0	0	0.08	0.30
Triplets	77	5	2	0	0	0	0	0.22	0.48

Tables 43 and 44 present CTT statistics and IRT parameter estimates for the 9 ATT items and the 9 VTT items of the MTT. Shown are the item means, standard deviations (SD), corrected item-total correlations (CITC), and SGRM item discrimination (*a*) and extremity parameters (b_1, b_2, b_3, and b_4).

Note that all of the corrected item-total correlations for the ATT component are large, with several approaching 0.6, and the IRT *a* parameter estimates for the slow part nearing 1.0 (excluding the 1.7 scaling factor). The b_4 parameter estimates are also noticeably higher for the medium and fast parts of the test reflecting the increases in difficulty associated with the higher

speeds of the target. With one exception, the *a* parameters were also quite a bit lower due to the difficult nature of the task.

Table 43. CTT and IRT Statistics for the 9 ATT Items of the MTT

ATT Item Name	Polytomous Responses			SGRM Parameters				
	Mean	SD	CITC	*a*	b_1	b_2	b_3	b_4
MTT_ATT_slow1p	2.24	1.30	.59	0.97	-1.55	-0.74	-.01	1.41
MTT_ATT_slow2p	2.39	1.34	.56	0.91	-1.81	-0.82	-.20	.98
MTT_ATT_slow3p	2.33	1.31	.59	0.97	-1.68	-0.81	-.06	1.13
MTT_ATT_med1p	1.72	1.34	.57	0.93	-1.03	-0.09	.66	1.77
MTT_ATT_med2p	1.92	1.33	.55	0.84	-1.27	-0.43	.43	1.81
MTT_ATT_med3p	1.94	1.31	.53	0.80	-1.52	-0.42	.50	1.75
MTT_ATT_fast1p	1.59	1.26	.43	0.59	-1.43	0.10	1.09	2.84
MTT_ATT_fast2p	1.52	1.20	.55	0.81	-1.12	0.11	1.03	2.74
MTT_ATT_fast3p	1.37	1.20	.44	0.60	-1.03	0.47	1.48	3.40

In Table 44, it can be seen that the corrected item-total correlations for the VTT component of the MTT are good, although smaller than those for the ATT component items. Most of the CITCs for VTT are in the 0.4 to 0.5 range, and the *a* parameters are generally in the 0.60s and 0.70s (excluding the 1.7 scaling factor). Interestingly, contrary to what was found for the VTT items in the ATTVTT, the b_3 and b_4 parameters illustrated here show the predicted increases in extremity as the speed of the target increased. If this is an effect due to practice during the ATTVTT subtest, then the implication is that a short practice before ATTVTT might help familiarize examinees with the controls and requirements and thus increase the discriminating power of the items.

Table 44. CTT and IRT Statistics for the 9 VTT Items of the MTT

VTT Item Name	Polytomous Responses			SGRM Parameters				
	Mean	SD	CITC	*a*	b_1	b_2	b_3	b_4
MTT_VTT_slow1p	1.85	1.26	.43	0.63	-2.05	-0.19	.95	1.99
MTT_VTT_slow2p	1.74	1.31	.48	0.75	-1.39	-0.09	.89	1.83
MTT_VTT_slow3p	1.46	1.25	.47	0.72	-1.07	0.37	1.24	2.48
MTT_VTT_med1p	1.17	1.06	.50	0.76	-0.87	0.74	2.07	3.03
MTT_VTT_med2p	1.25	1.05	.42	0.60	-1.27	0.82	2.21	3.57
MTT_VTT_med3p	1.29	1.06	.49	0.75	-1.04	0.37	1.90	3.31
MTT_VTT_fast1p	1.09	.95	.44	0.63	-0.89	0.90	2.67	5.06
MTT_VTT_fast2p	1.02	.96	.49	0.79	-0.67	1.01	2.31	3.75
MTT_VTT_fast3p	1.10	.99	.47	0.71	-0.83	0.81	2.33	3.80

Figures 17 and 18 show the test information functions for the ATT and VTT components of the MTT. Both plots confirm that the test is informative over a wide range of trait levels, but measurement precision declines somewhat at very low thetas because of the difficult nature of the task. Interestingly, as shown in Figure 18, the VTT items combine to provide more information at the upper end of the trait continuum than the ATT items. Research is needed to determine whether the improvements in MTT VTT item discrimination and, thus, information stem from practice (increased automaticity) due to having just completed ATTVTT.

Figure 17. Test Information Function for the 9 ATT Items of the MTT

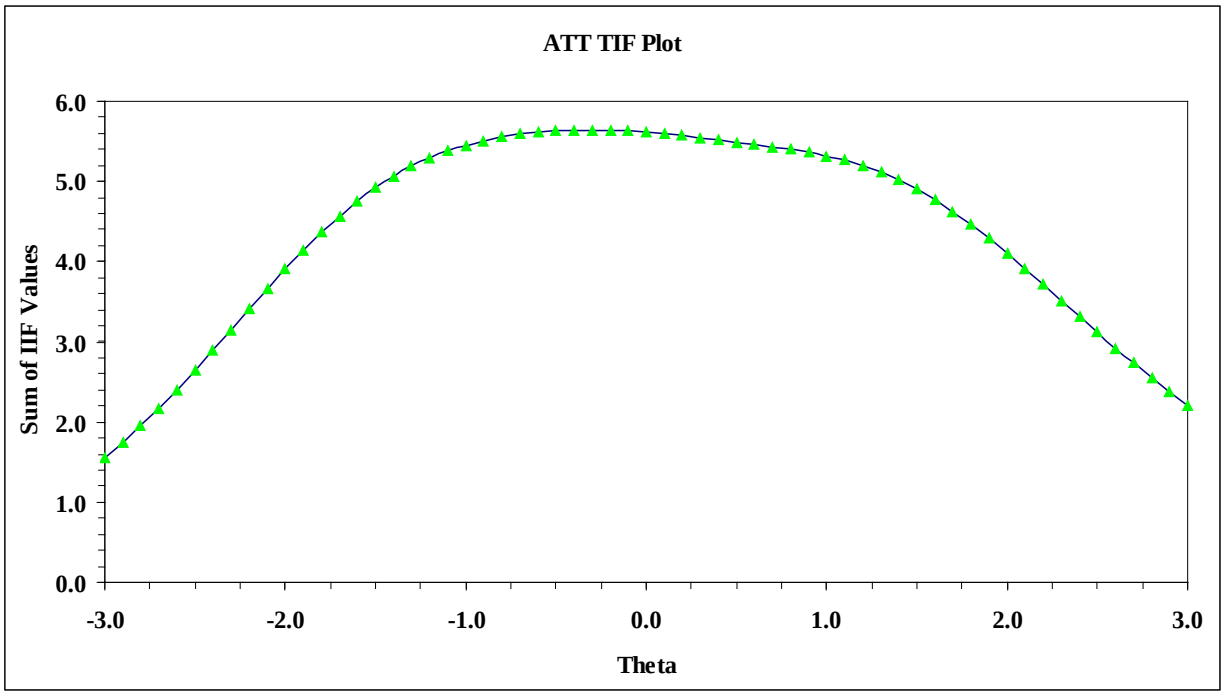
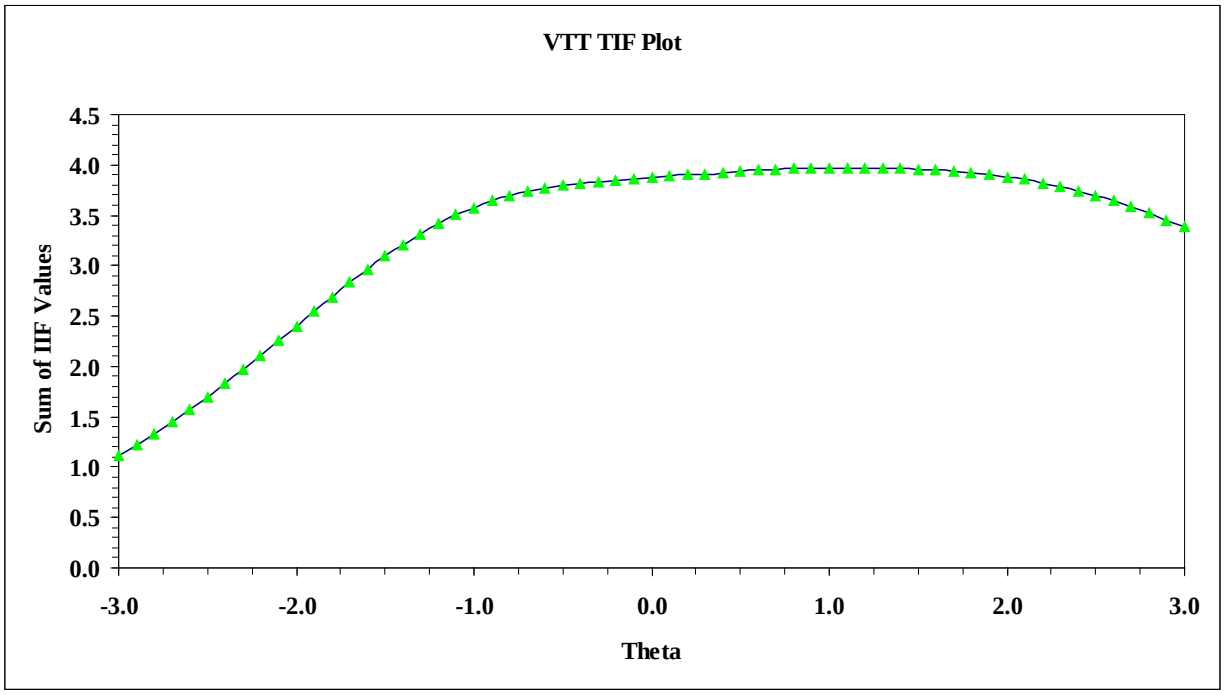


Figure 18. Test Information Function for the 9 VTT Items of the MTT



MTT Scale Scores

The subtest level indicators for the ATT, VTT, and DLT components of the MTT were also examined separately for SPs, SNFOs, and the total sample. The total numbers\ of redirects (MTT ATT Redirects and MTT VTT Redirects), the average distance between the respective crosshairs and the targets during the test (MTT ATT Average Distance and MTT VTT Average Distance), the total numbers of on-target responses (MTT ATT Total On Target and MTT VTT Total On Target), the IRT score (MTT ATT IRT Score and MTT VTT IRT Score), and the DLT total correct score (MTT DLT Total Correct) are all indicators of examinee ability. As with the individually administered ATT and VTT assessments, the average distance measures are negatively related to the scores for the other components in the MTT. The ATT and VTT components themselves correlated in the 0.40s and the correlations of both with DLT component scores were in the mid 0.10s and mid 0.20s (see Table 48).

Tables 45 and 46 show descriptive statistics for the ATT and VTT components of the MTT when analyzed separately using the total sample, SPs, and SNFOs. The SPs performed better on the ATT component by about a half of the total sample SD, but the effect size differences for VTT components were quite small. There were virtually no differences in terms of DLT Total Correct (see Table 47).

Table 45. Performance on the ATT Component of the MTT Across SP and SNFO Student Groups

Program	N	MTT ATT Redirects		MTT ATT Average Distance		MTT ATT Total On Target		MTT ATT IRT Score	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
SNFO	89	12.66	7.74	120.57	40.51	45.99	27.06	-0.30	0.90
SP	310	16.48	8.18	103.61	35.20	58.81	28.04	0.08	0.90
Total	399	15.62	8.23	107.40	37.08	55.94	28.30	0.00	0.91

Table 46. Performance on the VTT Component of the MTT Across SP and SNFO Student Groups

Program	N	MTT VTT Redirects		MTT VTT Average Distance		MTT VTT Total On Target		MTT VTT IRT Score	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
SNFO	89	15.62	6.33	104.97	30.32	74.42	29.13	-0.06	0.82
SP	310	16.81	7.22	98.61	30.02	80.45	33.36	0.02	0.89
Total	399	16.54	7.04	100.03	30.17	79.10	32.52	0.00	0.88

Table 47. Performance on the DLT Component of the MTT Across SP and SNFO Student Groups

Program	N	MTT DLT Total Correct	
		Mean	SD
SNFO	89	19.21	7.22
SP	308	19.70	7.17
Total	399	19.59	7.17

Table 48. Correlations Between the DLT, ATT, and VTT Component Scores of the MTT

	MTT ATT Redirects	MTT ATT Average Distance	MTT ATT Total On Target	MTT ATT IRT Score	MTT VTT Redirects	MTT VTT Average Distance	MTT VTT Total On Target	MTT VTT IRT Score	MTT DLT Total Correct
MTT ATT Redirects	1	-.870**	.992**	.920**	.459**	-.478**	.458**	.415**	.163**
MTT ATT Average Distance	-.870**	1	-.866**	-.871**	-.339**	.438**	-.341**	-.335**	-.187**
MTT ATT Total On Target	.992**	-.866**	1	.929**	.460**	-.477**	.460**	.416**	.174**
MTT ATT IRT Score	.920**	-.871**	.929**	1	.382**	-.427**	.384**	.385**	.161**
MTT VTT Redirects	.459**	-.339**	.460**	.382**	1	-.898**	.993**	.898**	.205**
MTT VTT Average Distance	-.478**	.438**	-.477**	-.427**	-.898**	1	-.896**	-.857**	-.233**
MTT VTT Total On Target	.458**	-.341**	.460**	.384**	.993**	-.896**	1	.902**	.206**
MTT VTT IRT Score	.415**	-.335**	.416**	.385**	.898**	-.857**	.902**	1	.207**
MTT DLT Total Correct	.163**	-.187**	.174**	.161**	.205**	-.233**	.206**	.207**	1

Table 49 shows the correlations of the scores for the ATT and DLT subcomponents of the MTT with other potential predictors of training performance. ATT component scores showed correlations in the 0.2 to 0.3 range with most criteria and a correlation of 0.42 with simExperience. As with the results for ATTVTT, the correlations with aTraining and Education were at or below 0.1. Correlations of DLT Total Correct with aTraining, Education, simExperience, and flightHours were similarly low, but the correlations were in the 0.20s for MCT_Post2004, AQR_Post2004, and OAR_Post2004.

Table 49. Correlations Between the ATT and DLT Component Scores of the MTT and Other Predictors

	N	Mean	SD	MTT ATT Redirects	MTT ATT Average Distance	MTT ATT Total On Target	MTT ATT IRT Score	MTT DLT Total Correct
aTraining	388	0.23	0.72	.094	-.072	.091	.099	.008
Education	383	2.88	0.59	.102*	-.083	.098	.100*	.165**
simExperience	397	0.74	0.83	.420**	-.339**	.421**	.396**	.078
flightHours	389	0.69	1.38	.131**	-.085	.135**	.133**	.003
ANI_RAW	330	0.58	0.53	.195**	-.204**	.195**	.198**	-.025
MST_RAW	330	0.34	0.67	.115*	-.102	.122*	.128*	.192**
RCT_RAW	330	0.43	0.53	.040	-.036	.031	.038	.172**
SAT_Post2004	330	0.76	0.64	.211**	-.194**	.202**	.213**	.089
MCT_Post2004	330	0.50	0.64	.211**	-.198**	.219**	.215**	.216**
AQR_Post2004	330	0.55	0.52	.271**	-.260**	.275**	.279**	.204**
PFAR_Post2004	330	0.67	0.50	.292**	-.286**	.292**	.298**	.118*
FOFAR_Post2004	330	0.65	0.53	.254**	-.239**	.252**	.264**	.186**
OAR_Post2004	330	0.50	0.62	.206**	-.190**	.213**	.213**	.250**

Table 50 shows correlations of the MTT VTT component scores with other potential predictors of training performance. Unlike with ATTVTT, the VTT component scores for MTT showed moderate correlations (several in the 0.30s) with other performance predictors –most notably MTT VTT Average Distance with AQR_, PFAR_, and FOFAR_. The correlations with simExperience were 0.25 to 0.33 in magnitude, with the IRT scores showing the lowest (0.255), possibly due to loss of information through polytomization of the continuous data

Table 50. Correlations Between the VTT Component Scores of the MTT and Other Predictors

	N	Mean	SD	MTT VTT Redirects	MTT VTT Average Distance	MTT VTT Total On Target	MTT VTT IRT Score
aTraining	388	0.23	0.72	.054	-.039	.047	.041
Education	383	2.88	0.59	-.003	-.024	-.006	.032
simExperience	397	0.74	0.83	.326**	-.311**	.330**	.255**
flightHours	389	0.69	1.38	.097	-.020	.091	.049
ANI_RAW	330	0.58	0.53	.143**	-.176**	.133*	.114*
MST_RAW	330	0.34	0.67	.207**	-.213**	.199**	.189**
RCT_RAW	330	0.43	0.53	.139*	-.142**	.151**	.153**
SAT_Post2004	330	0.76	0.64	.209**	-.251**	.216**	.208**
MCT_Post2004	330	0.50	0.64	.197**	-.225**	.202**	.187**
AQR_Post2004	330	0.55	0.52	.287**	-.323**	.283**	.263**
PFAR_Post2004	330	0.67	0.50	.263**	-.311**	.260**	.239**
FOFAR_Post2004	330	0.65	0.53	.309**	-.345**	.306**	.289**
OAR_Post2004	330	0.50	0.62	.239**	-.262**	.240**	.225**

Tables 51, 52, 53, and 54 present the correlations between the ATT, VTT, and DLT component scores of the MTT and training criteria (block grades and training composites) for the total sample and for student pilots only. The results for the SP sample shown in Table 52 indicate that the ATT component correlations were highest with I20, Instruments_ALL, and Instruments_BASIC, with values in the mid to high 0.20s; the same correlations were smaller in the total sample.

The VTT results for SPs in Table 54 revealed sizeable correlations with I20, I21, F40, F42, Instruments _BASIC, and Formation_Aircraft, as might have been expected based on the good CITCs, discrimination, and extremity parameters shown in Table 44. Overall, both the VTT and ATT components have correlations with criteria large enough to be of practical importance for selection.

The DLT Total Correct correlations (Tables 51 and 52, last column) were generally smaller than those for the ATT and VTT components (most were in the 0.10s). However, they should provide incremental validity for selection because of the correlations with Instruments_ and Navigation_ criteria in the 0.15 to 0.20 range and the low correlations with the ATT and VTT component scores. Note also that the ATT measures had near zero correlations with Navigation_AIRCRAFT but DLT Total Correct had a substantial (0.20) correlation with the training criterion.

Table 51. Correlations Between the ATT and DLT Component Scores of the MTT and Navy Pilot Training Criteria for the Total Sample

		Total Sample				
Training Block Name	N	MTT ATT Redirects	MTT ATT Average Distance	MTT ATT Total On Target	MTT ATT IRT Score	MTT DLT Total Correct
C20	398	.077	-.089	.077	.074	.142**
C40	86	-.013	.008	-.017	-.020	.003
C41	373	.098	-.037	.094	.071	.000
C42	366	.061	-.033	.059	.065	.056
C43	270	.097	-.074	.101	.104	.109
C45	262	.116	-.136*	.116	.149*	.039
C46	246	.178**	-.207**	.178**	.156*	.017
C47	238	.111	-.055	.107	.104	-.005
I20	386	.239**	-.206**	.242**	.229**	.149**
I21	299	.220**	-.232**	.237**	.223**	.127*
I22	207	.088	-.142*	.093	.123	.106
I23	206	.126	-.134	.119	.132	.147*
I24	194	.198**	-.174*	.187**	.196**	.100
I25	188	.135	-.129	.123	.151*	.055
I40	379	.113*	-.132**	.117*	.112*	.125*
I41	282	.230**	-.234**	.232**	.221**	.061
I42	233	.165*	-.157*	.168*	.162*	.112
I43	227	.097	-.111	.108	.091	.060
F40	225	.173**	-.174**	.175**	.149*	.127
F42	209	.231**	-.236**	.228**	.188**	.061
N40	183	.096	-.063	.097	.115	.136
N41	181	-.024	.018	-.020	.007	.168*
Contact_Simulation	398	.077	-.089	.077	.074	.142**
Contact_AIRCRAFT	377	.124*	-.091	.122*	.120*	.040
Contact_ALL	398	.122*	-.110*	.120*	.124*	.101*
Instruments_Simulation	386	.205**	-.188**	.207**	.203**	.159**
Instruments_AIRCRAFT	379	.172**	-.176**	.174**	.151**	.147**
Instruments_ALL	386	.216**	-.212**	.217**	.208**	.166**
Instruments_BASIC	386	.212**	-.203**	.218**	.205**	.164**
Instruments_RADIO	289	.183**	-.217**	.189**	.196**	.090
Instruments_NAVIGATION	244	.190**	-.194**	.189**	.205**	.147*
Navigation_AIRCRAFT	183	.043	-.020	.045	.069	.200**
Formation_AIRCRAFT	225	.198**	-.208**	.199**	.166*	.117

Navy Standard Score (NSS)	398	.191**	-.195**	.196**	.182**	.154**

Table 52. Correlations Between the ATT and DLT Component Scores of the MTT and Navy Pilot Training Criteria for the Student Pilots Only

	Students Pilots (SPs)					
Training Block Name	N	MTT ATT Redirects	MTT ATT Average Distance	MTT ATT Total On Target	MTT ATT IRT Score	MTT DLT Total Correct
C20	309	.101	-.115*	.101	.102	.135*
C40	-	-	-	-	-	-
C41	291	.146*	-.095	.142*	.106	.039
C42	283	.115	-.086	.107	.114	.074
C43	270	.097	-.074	.101	.104	.109
C45	262	.116	-.136*	.116	.149*	.039
C46	246	.178**	-.207**	.178**	.156*	.017
C47	238	.111	-.055	.107	.104	-.005
I20	302	.288**	-.274**	.290**	.280**	.135*
I21	299	.220**	-.232**	.237**	.223**	.127*
I22	207	.088	-.142*	.093	.123	.106
I23	206	.126	-.134	.119	.132	.147*
I24	194	.198**	-.174*	.187**	.196**	.100
I25	188	.135	-.129	.123	.151*	.055
I40	297	.138*	-.150**	.134*	.138*	.132*
I41	200	.224**	-.207**	.215**	.199**	.076
I42	183	.185*	-.162*	.181*	.180*	.107
I43	178	.066	-.060	.071	.058	.029
F40	225	.173**	-.174**	.175**	.149*	.127
F42	209	.231**	-.236**	.228**	.188**	.061
N40	183	.096	-.063	.097	.115	.136
N41	181	-.024	.018	-.020	.007	.168*
Contact_Simulation	309	.101	-.115*	.101	.102	.135*
Contact_AIRCRAFT	291	.187**	-.152**	.183**	.182**	.058
Contact_ALL	309	.164**	-.156**	.162**	.170**	.115*
Instruments_Simulation	302	.252**	-.259**	.251**	.254**	.145*
Instruments_AIRCRAFT	297	.212**	-.208**	.207**	.188**	.150**
Instruments_ALL	302	.261**	-.265**	.257**	.255**	.159**
Instruments_BASIC	302	.258**	-.260**	.261**	.255**	.158**

Instruments_RADIO	207	.161*	-.185**	.158*	.167*	.121
Instruments_NAVIGATION	194	.195**	-.182*	.185**	.214**	.144*
Navigation_AIRCRAFT	183	.043	-.020	.045	.069	.200**
Formation_AIRCRAFT	225	.198**	-.208**	.199**	.166*	.117
Navy Standard Score (NSS)	309	.235**	-.240**	.233**	.224**	.158**

Table 53. Correlations Between the VTT Component Scores of the MTT and Navy Pilot Training Criteria for the Total Sample

	Total Sample				
Training Block Name	N	MTT VTT Redirects	MTT VTT Average Distance	MTT VTT Total On Target	MTT VTT IRT Score
C20	397	.045	-.032	.042	.028
C40	86	.219*	-.200	.204	.277**
C41	372	.111*	-.086	.112*	.070
C42	365	.106*	-.094	.107*	.080
C43	269	.074	-.069	.072	.080
C45	261	.131*	-.125*	.143*	.118
C46	245	.103	-.117	.103	.077
C47	237	.147*	-.138*	.139*	.118
I20	385	.280**	-.274**	.273**	.256**
I21	298	.293**	-.293**	.287**	.268**
I22	206	.065	-.112	.060	.059
I23	205	.090	-.145*	.092	.046
I24	193	.177*	-.184*	.176*	.151*
I25	187	.089	-.108	.085	.077
I40	378	.219**	-.195**	.217**	.165**
I41	281	.140*	-.203**	.136*	.108
I42	232	.113	-.126	.113	.102
I43	226	.092	-.104	.097	.074
F40	224	.225**	-.263**	.228**	.218**
F42	208	.205**	-.234**	.197**	.209**
N40	182	.104	-.057	.110	.104
N41	180	.083	-.076	.088	.095
Contact_Simulation	397	.045	-.032	.042	.028
Contact_AIRCRAFT	376	.184**	-.167**	.182**	.171**
Contact_ALL	397	.136**	-.119*	.132**	.124*
Instruments_Simulation	385	.242**	-.245**	.237**	.231**
Instruments_AIRCRAFT	378	.195**	-.186**	.192**	.141**

Instruments_ALL	385	.253**	-.256**	.247**	.223**
Instruments_BASIC	385	.308**	-.293**	.303**	.269**
Instruments_RADIO	288	.111	-.182**	.106	.082
Instruments_NAVIGATION	243	.155*	-.176**	.154*	.133*
Navigation_AIRCRAFT	182	.103	-.077	.109	.102
Formation_AIRCRAFT	224	.223**	-.277**	.223**	.226**
Navy Standard Score (NSS)	397	.212**	-.214**	.208**	.181**

Table 54. Correlations Between the VTT Component Scores of the MTT and Navy Pilot Training Criteria for the Students Sample Only

	Students Pilots (SPs)				
Training Block Name	N	MTT VTT Redirects	MTT VTT Average Distance	MTT VTT Total On Target	MTT VTT IRT Score
C20	308	.041	-.023	.040	.024
C40	-	-	-	-	-
C41	290	.170**	-.135*	.169**	.120*
C42	282	.171**	-.163**	.170**	.131*
C43	269	.074	-.069	.072	.080
C45	261	.131*	-.125*	.143*	.118
C46	245	.103	-.117	.103	.077
C47	237	.147*	-.138*	.139*	.118
I20	301	.300**	-.306**	.288**	.265**
I21	298	.293**	-.293**	.287**	.268**
I22	206	.065	-.112	.060	.059
I23	205	.090	-.145*	.092	.046
I24	193	.177*	-.184*	.176*	.151*
I25	187	.089	-.108	.085	.077
I40	296	.224**	-.190**	.223**	.169**
I41	199	.145*	-.189**	.142*	.111
I42	182	.108	-.115	.109	.086
I43	177	.096	-.077	.098	.073
F40	224	.225**	-.263**	.228**	.218**
F42	208	.205**	-.234**	.197**	.209**
N40	182	.104	-.057	.110	.104
N41	180	.083	-.076	.088	.095
Contact_Simulation	308	.041	-.023	.040	.024
Contact_AIRCRAFT	290	.199**	-.180**	.199**	.163**
Contact_ALL	308	.134*	-.113*	.132*	.112*

Instruments_Simulation	301	.256**	-.272**	.246**	.234**
Instruments_AIRCRAFT	296	.203**	-.184**	.199**	.143*
Instruments_ALL	301	.264**	-.270**	.255**	.228**
Instruments_BASIC	301	.328**	-.318**	.320**	.282**
Instruments_RADIO	206	.107	-.160*	.104	.075
Instruments_NAVIGATION	193	.163*	-.167*	.160*	.132
Navigation_AIRCRAFT	182	.103	-.077	.109	.102
Formation_AIRCRAFT	224	.223**	-.277**	.223**	.226**
Navy Standard Score (NSS)	308	.224**	-.220**	.218**	.183**

Table 55 presents the results of multiple regression analyses using the ATT, VTT, and DLT components of the MTT as predictors of composite performance criteria for the student pilot sample. Prior to these analyses, we checked for the interaction between ATT and VTT components by conducted a series of moderated regression analyses (not shown here). In the 20 analyses conducted, no significant interactions were found, so we did not include interaction terms in the models presented in Table 55.

Similar to the ATTVTT subtest, the ATT component scores of the MTT subtest significantly predicted three of the five training performance criteria. The exceptions were Navigation and Formation grades, which were predicted by DLT Total Correct scores and the VTT component scores respectively. Notably, for predicting overall NSS grades, all three predictors appeared to be useful with analyses showing several to be statistically significant. So, unlike the ATTVTT where the ATT subcomponent scores were the only significant predictors of training performance, the VTT and DLT component scores of the MTT provided incremental validities beyond the ATT in many analyses and merit attention.

Table 55. MTT Multiple Regression Results using the DLT, ATT and VTT Component Scores as Predictors of Navy Pilot Training Criteria

	Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	R
		B	Std. Error	Beta			
Contact ALL	(Constant)	44.76	1.45		30.81	0.00	0.198
	MTT ATT Redirects	0.11	0.06	0.13	1.98	0.05	
	MTT VTT Redirects	0.06	0.06	0.06	0.87	0.38	
	MTT DLT Total Correct	0.09	0.06	0.09	1.53	0.13	
	(Constant)	51.27	2.23		22.99	0.00	0.188
	MTT ATT Average Distance	-0.03	0.01	-0.13	-2.06	0.04	
	MTT VTT Average Distance	-0.01	0.02	-0.04	-0.62	0.54	
	MTT DLT Total Correct	0.09	0.06	0.09	1.46	0.14	
	(Constant)	44.69	1.48		30.26	0.00	0.196
	MTT ATT Total On Target	0.03	0.02	0.12	1.95	0.05	

	MTT VTT Total On Target	0.01	0.01	0.06	0.88	0.38	
	MTT DLT Total Correct	0.09	0.06	0.09	1.53	0.13	
	(Constant)	47.48	1.22		38.77	0.00	0.184
	MTT ATT IRT Score	1.06	0.50	0.13	2.12	0.04	
	MTT VTT IRT Score	0.31	0.50	0.04	0.61	0.54	
	MTT DLT Total Correct	0.09	0.06	0.08	1.47	0.14	
Instruments ALL	(Constant)	40.82	1.61		25.36	0.00	0.322
	MTT ATT Redirects	0.17	0.06	0.17	2.70	0.01	
	MTT VTT Redirects	0.19	0.07	0.17	2.65	0.01	
	MTT DLT Total Correct	0.12	0.06	0.10	1.79	0.07	
	(Constant)	56.15	2.45		22.90	0.00	0.329
	MTT ATT Average Distance	-0.04	0.01	-0.18	-2.91	0.00	
	MTT VTT Average Distance	-0.05	0.02	-0.18	-2.91	0.00	
	MTT DLT Total Correct	0.10	0.07	0.08	1.48	0.14	
	(Constant)	40.73	1.64		24.85	0.00	0.316
	MTT ATT Total On Target	0.05	0.02	0.17	2.69	0.01	
	MTT VTT Total On Target	0.04	0.02	0.16	2.52	0.01	
	MTT DLT Total Correct	0.12	0.06	0.10	1.80	0.07	
	(Constant)	46.54	1.36		34.20	0.00	0.296
	MTT ATT IRT Score	1.61	0.56	0.18	2.90	0.00	
	MTT VTT IRT Score	1.23	0.56	0.13	2.19	0.03	
	MTT DLT Total Correct	0.12	0.07	0.10	1.79	0.07	
Navigation AIRCRAFT	(Constant)	44.18	2.39		18.46	0.00	0.210
	MTT ATT Redirects	-0.03	0.08	-0.03	-0.31	0.75	
	MTT VTT Redirects	0.09	0.09	0.08	0.94	0.35	
	MTT DLT Total Correct	0.25	0.10	0.19	2.50	0.01	
	(Constant)	45.06	3.71		12.13	0.00	0.204
	MTT ATT Average Distance	0.01	0.02	0.04	0.54	0.59	
	MTT VTT Average Distance	-0.01	0.02	-0.04	-0.48	0.63	
	MTT DLT Total Correct	0.26	0.10	0.20	2.56	0.01	
	(Constant)	44.05	2.42		18.18	0.00	0.212
	MTT ATT Total On Target	-0.01	0.02	-0.03	-0.32	0.75	
	MTT VTT Total On Target	0.02	0.02	0.08	1.03	0.31	

	MTT DLT Total Correct	0.25	0.10	0.19	2.49	0.01	
Formation AIRCRAFT	(Constant)	45.33	2.12		21.38	0.00	0.210
	MTT ATT IRT Score	0.07	0.72	0.01	0.10	0.92	
	MTT VTT IRT Score	0.61	0.74	0.06	0.82	0.41	
	MTT DLT Total Correct	0.25	0.10	0.19	2.47	0.01	
	(Constant)	42.72	2.17		19.70	0.00	0.263
	MTT ATT Redirects	0.13	0.08	0.12	1.71	0.09	
	MTT VTT Redirects	0.20	0.09	0.16	2.26	0.03	
	MTT DLT Total Correct	0.09	0.09	0.07	1.07	0.28	
	(Constant)	58.54	3.23		18.12	0.00	0.300
	MTT ATT Average Distance	-0.03	0.02	-0.11	-1.60	0.11	
	MTT VTT Average Distance	-0.07	0.02	-0.22	-3.12	0.00	
	MTT DLT Total Correct	0.05	0.09	0.04	0.60	0.55	
	(Constant)	42.50	2.20		19.31	0.00	0.264
	MTT ATT Total On Target	0.04	0.02	0.12	1.74	0.08	
	MTT VTT Total On Target	0.04	0.02	0.16	2.25	0.03	
	MTT DLT Total Correct	0.09	0.09	0.07	1.06	0.29	
	(Constant)	48.18	1.83		26.36	0.00	0.254
	MTT ATT IRT Score	0.86	0.67	0.09	1.29	0.20	
	MTT VTT IRT Score	1.88	0.71	0.18	2.63	0.01	
	MTT DLT Total Correct	0.09	0.09	0.07	1.02	0.31	
Navy Standard Score (NSS)	(Constant)	40.45	1.89		21.44	0.00	0.290
	MTT ATT Redirects	0.19	0.07	0.16	2.57	0.01	
	MTT VTT Redirects	0.17	0.08	0.13	2.03	0.04	
	MTT DLT Total Correct	0.15	0.08	0.11	1.96	0.05	
	(Constant)	55.61	2.89		19.26	0.00	0.292
	MTT ATT Average Distance	-0.05	0.02	-0.17	-2.85	0.00	
	MTT VTT Average Distance	-0.04	0.02	-0.12	-2.04	0.04	
	MTT DLT Total Correct	0.13	0.08	0.10	1.72	0.09	
	(Constant)	40.33	1.92		21.02	0.00	0.286
	MTT ATT Total On Target	0.06	0.02	0.16	2.59	0.01	
	MTT VTT Total On Target	0.03	0.02	0.12	1.93	0.05	
	MTT DLT Total Correct	0.15	0.08	0.11	1.96	0.05	
	(Constant)	46.18	1.59		28.99	0.00	0.266

MTT ATT IRT Score	1.79	0.65	0.16	2.73	0.01
MTT VTT IRT Score	1.06	0.66	0.10	1.62	0.11
MTT DLT Total Correct	0.15	0.08	0.11	1.94	0.05

In summary, the MTT provides valuable information for decision making. Both the ATT and VTT components showed high discrimination and a wide range of difficulties when analyzed using CTT and IRT methods. Both showed moderate correlations with criteria and appeared to contribute independently to the prediction of training grades. No evidence of nonadditive effects was found, as the interaction terms in the moderated multiple regression analyses were not significant.

DLT Total Correct scores also showed correlations near 0.2 with instruments and navigation criteria and low correlations with the other MTT components. The correlation with NSS Grades is particularly important, because the other PBM scores had low correlations with this criterion. When forming a test battery that will provide useful prediction of all these criterion measures, the DLT should be considered.

EMERGENCY SCENARIO TEST (EST)
SCORING STRATEGIES AND VALIDITIES

In the EST (also referred to as AttVttEst), examinees must respond to three emergency scenarios involving fuel flow, engine power, and propeller position while performing one- and two-dimensional tracking tasks. The proper responses to the emergency scenarios are indicated to an examinee before beginning the subtest by way of a detailed instructions screen. Because an examinee must memorize a set of corrective actions, recall each action quickly in response to an emergency notification, and manipulate buttons on a throttle to resolve the emergency within 20 seconds of a notification (all the while performing VTT and ATT), the EST appears to measure a combination of abilities: general cognitive ability, spatial ability, psychomotor dexterity, and stress tolerance.

During the EST a variety of performance data are recorded: 1) dichotomously scored responses to the emergency scenarios with "1" being assigned if corrective actions are performed within the 4s time limit, and "0" being assigned if no action or the wrong actions are taken, 2) response time information when emergency scenarios are resolved successfully, and 3) distance information and numbers of redirects for the VTT and ATT components. Data for the EST, ATT, and VTT were analyzed separately using correlations, regressions, and IRT methods.

To explore the possibility that faster response times are more predictive of examinee training performance than slower response times, we rescored response time information for successfully resolved emergency scenarios as follows: Responses within 2.7 to 20.0 seconds = 1; responses within 2.4 to 2.69 seconds = 2; responses under 2.4 seconds = 3. These ordered polytomous data were used for IRT analysis based on SGRM. We also computed total emergency scenario scores by summing across the original dichotomous scores (EST Scenario Score) and across the constructed polytomous scores (EST Scores Poly) for higher-level analyses.

To form "item level" data for analyzing the ATT and VTT components separately using the SGRM IRT model, we sampled nine 7.2 second (18 400ms interval) time periods for each task and airplane speed. To reduce score dependencies between adjacent periods, we ignored data for one 400ms interval between each period, making the highest possible score for a time period 18, meaning that an examinee was on-target every time a measurement was taken. The VTT and ATT components of the EST last 120 seconds, or 40 seconds at each airplane speed.

Different thresholds were used to transform the continuous tracking data into 5-option polytomous responses for the IRT analyses. For the VTT component of the EST, the following categorization scheme was used: 0-1 = 0; 2-3 = 1; 4-5 = 2; 6-7 = 3, 8-18 = 4. For the ATT component, because very few examinees had on-target values larger than 6, a different scheme was used: 0 = 0; 1 = 1; 2 = 2; 3-4 = 3, 5-18 = 4. These thresholds were identical to those in the MTT subtest.

Item-Level CTT and IRT Analyses and Results for the ATT and VTT Components of the EST

Because the response data were scored polytomously with category codes of higher magnitude indicating better performance on the ATT and VTT components, SGRM (Samejima, 1969) for ordered polytomous responses was chosen for IRT analyses. To verify that the response data for each component were sufficiently unidimensional, we conducted separate principal component analyses of the ATT and VTT inter-item correlations. The scree plot for the ATT analysis is shown in Figure 19, and the scree plot for the VTT analysis is shown in Figure 20. In both cases, the data exhibited a strong first factor with the ratio of first to second eigenvalues exceeding 3.0, as recommended for application of a unidimensional IRT model (Drasgow & Parsons, 1983; Lord, 1980).

Figure 19. Scree Plot for the Principal Axis Factor Analysis of the 9 ATT Items of the EST

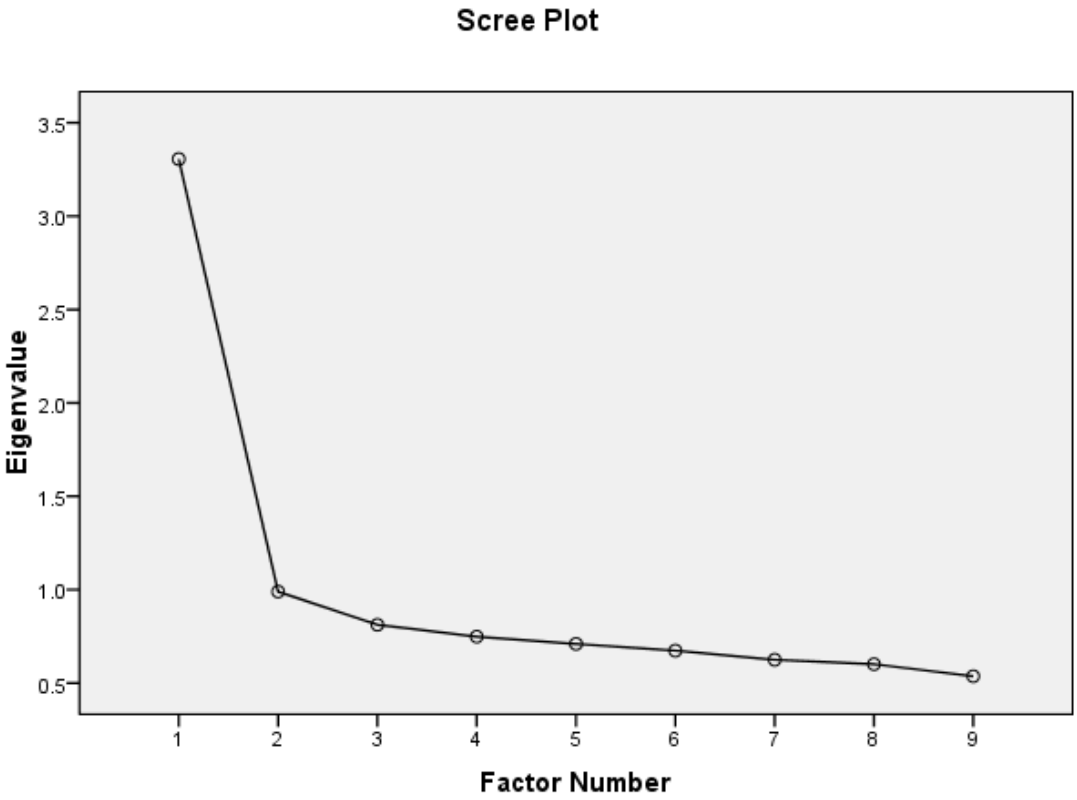
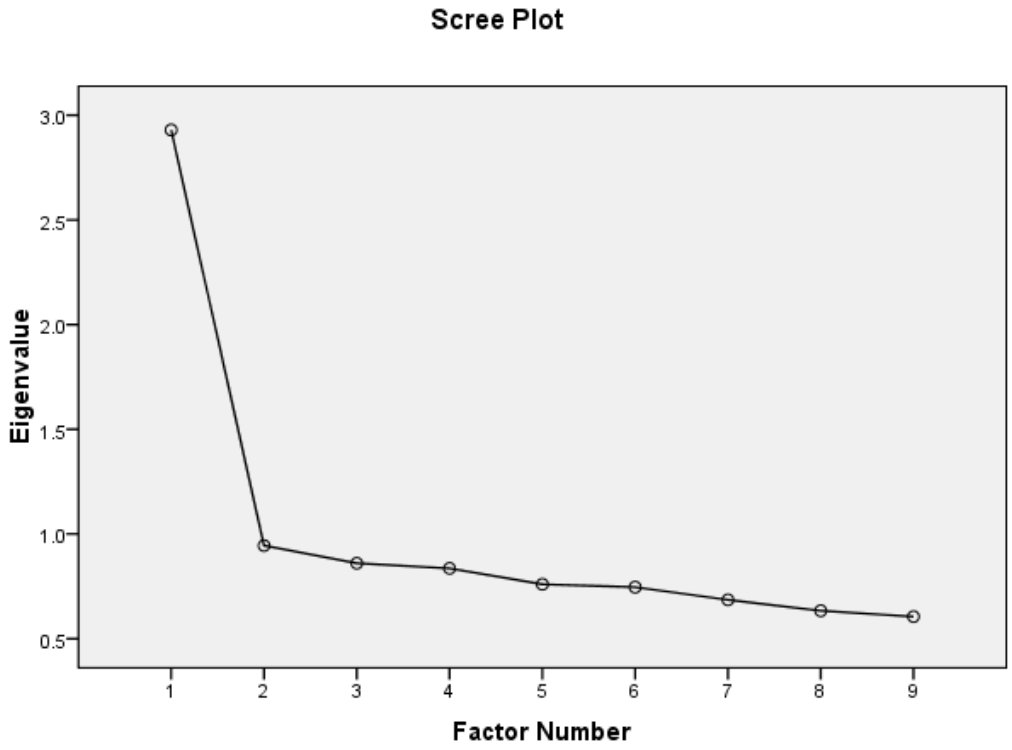


Figure 20. Scree Plot for the Principal Axis Factor Analysis of the 9 VTT Items of the EST



IRT Calibrations of the 9 ATT and 9 VTT Items of the EST

SGRM item parameters for the ATT and VTT components of the EST were estimated separately using the MULTILOG (Thissen, 1991) computer program (The command files were similar to those shown in previous sections of this report.). Because the data for each component were coded such that the responses to items fell within one of five ordered categories, there were five SGRM parameters to estimate per item: one discrimination parameter, *a*, and four extremity parameters, b_1, b_2, b_3, and b_4. Scoring and model-data fit analyses were performed using the MODFIT-Z 2.0 computer program (Stark, 2007). Separate parameter estimates, model-data fit statistics, and information functions are reported for the ATT and VTT components of the EST in the tables that follow.

Overall the fit plots and chi-square statistics indicated that SGRM fit the data for both the ATT and VTT components of the EST very well. As shown in Tables 56 and 57, the chi-square statistics were well below the threshold of 3, indicating good fit.

Table 56. Chi-Square Model-Data Fit Statistics for Items Created from the ATT Component Data of the EST

FREQUENCY TABLE OF ADJUSTED (N=3000) CHISQUARE/DF RATIOS									
	<1	1<2	2<3	3<4	4<5	5<7	>7	Mean	SD
Singlets	9	0	0	0	0	0	0	0	0
Doublets	35	0	1	0	0	0	0	0.111	0.476
Triplets	71	10	2	1	0	0	0	0.412	0.734

Table 57. Chi-Square Model-Data Fit Statistics for Items Created from the VTT Component Data of the EST

FREQUENCY TABLE OF ADJUSTED (N=3000) CHISQUARE/DF RATIOS									
	<1	1<2	2<3	3<4	4<5	5<7	>7	Mean	SD
Singlets	9	0	0	0	0	0	0	0	0
Doublets	34	1	1	0	0	0	0	0.157	0.532
Triplets	67	11	5	0	0	1	0	0.466	0.930

Tables 58 and 59 present CTT statistics and IRT parameter estimates for the 9 ATT items and the 9 VTT items of the EST. Shown are the item means, standard deviations (SD), corrected item-total correlations (CITC), and SGRM item discrimination (*a*) and extremity parameters (b_1, b_2, b_3, and b_4).

Note that all of the corrected item-total correlations for the ATT component are large, with several around 0.6, and the IRT *a* parameter estimates for the slow part varying from approximately 0.8 to 0.9 (excluding the 1.7 scaling factor). The b_4 parameter estimates are also noticeably higher for the medium and fast parts of the test reflecting the increased difficulty associated with higher speeds of the target. The *a* parameters are also quite a bit lower, perhaps due to the difficult nature of the task.

Table 58. CTT and IRT Statistics for the 9 ATT Items of the EST

ATT Item Name	Polytomous Responses			SGRM Parameters				
	Mean	SD	CITC	*a*	b_1	b_2	b_3	b_4
EST_ATT_slow1p	1.95	1.35	.61	0.88	-1.30	-0.46	.37	1.61
EST_ATT_slow2p	2.51	1.35	.63	0.93	-1.86	-0.90	-.33	.80
EST_ATT_slow3p	1.89	1.33	.60	0.81	-1.37	-0.29	.47	1.83
EST_ATT_med1p	1.70	1.32	.55	0.68	-1.28	-0.08	.74	2.43
EST_ATT_med2p	1.55	1.34	.41	0.49	-1.10	-0.01	1.14	3.19
EST_ATT_med3p	1.32	1.23	.46	0.53	-0.92	0.61	1.63	3.55
EST_ATT_fast1p	1.35	1.18	.51	0.65	-0.97	0.36	1.46	3.30
EST_ATT_fast2p	1.44	1.22	.51	0.64	-1.05	0.26	1.35	3.09
EST_ATT_fast3p	1.35	1.26	.52	0.66	-0.77	0.40	1.31	2.97

In Table 59, it can be seen that the corrected item-total correlations for the VTT component of EST are also good, although slightly smaller than those for the ATT component items. All of the CITCs for VTT are in the 0.4 to 0.5 range and the *a* parameters ranged from 0.5 to 0.8 (excluding the 1.7 scaling factor). Similar to the VTT results in the MTT, the b_4 parameters here are noticeably larger for the portions of the test involving a fast moving target.

Table 59. CTT and IRT Statistics for the 9 VTT Items of the EST

VTT Item Name	Polytomous Responses			SGRM Parameters				
	Mean	SD	CITC	*a*	b_1	b_2	b_3	b_4
EST_VTT_slow1p	1.59	1.33	.51	0.65	-1.28	0.22	1.18	2.07
EST_VTT_slow2p	1.70	1.34	.59	0.81	-1.12	-0.06	.88	1.77
EST_VTT_slow3p	1.35	1.29	.54	0.71	-0.69	0.38	1.37	2.44
EST_VTT_med1p	1.08	1.10	.55	0.77	-0.48	0.80	2.07	3.13
EST_VTT_med2p	1.08	1.07	.40	0.49	-0.82	1.20	2.84	4.36
EST_VTT_med3p	.95	1.03	.48	0.64	-0.36	1.21	2.55	3.95
EST_VTT_fast1p	1.09	1.02	.41	0.50	-0.97	1.20	2.88	4.61
EST_VTT_fast2p	1.15	1.05	.49	0.65	-0.84	0.72	2.27	3.71
EST_VTT_fast3p	1.08	1.06	.44	0.53	-0.73	1.02	2.69	4.33

Figures 21 and 22 show the test information functions for the ATT and VTT components of the EST. Both plots confirm that the test is informative over a wide range of trait levels, but measurement precision declines somewhat at very low thetas because of the difficult nature of the task.

Figure 21. Test Information Function for the 9 ATT Items of the EST

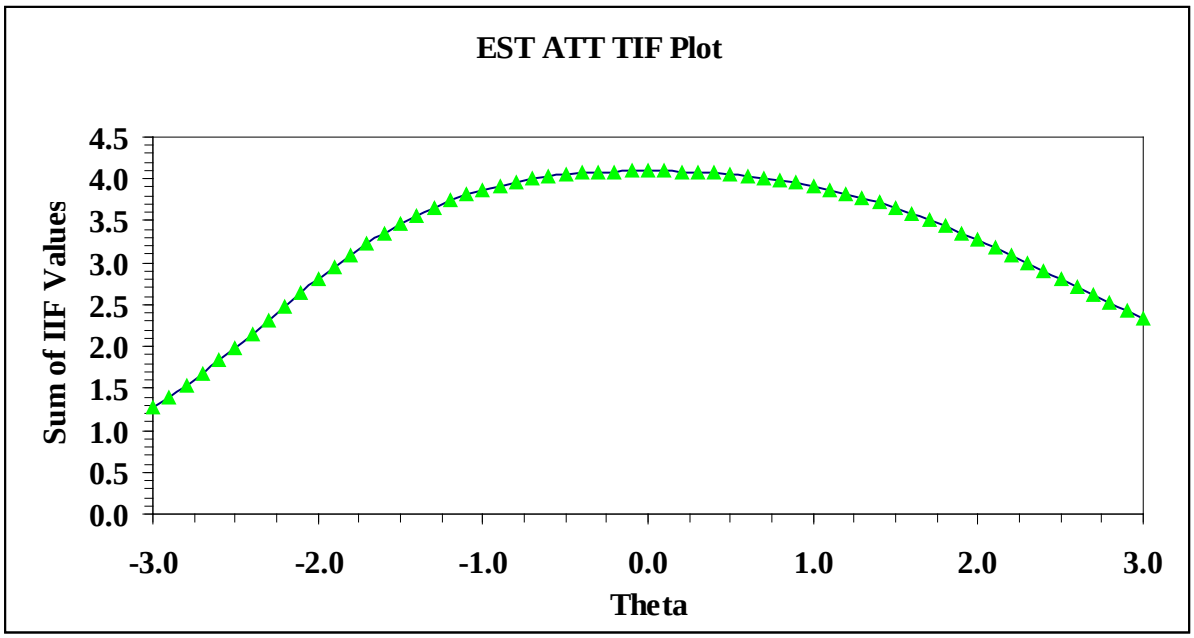
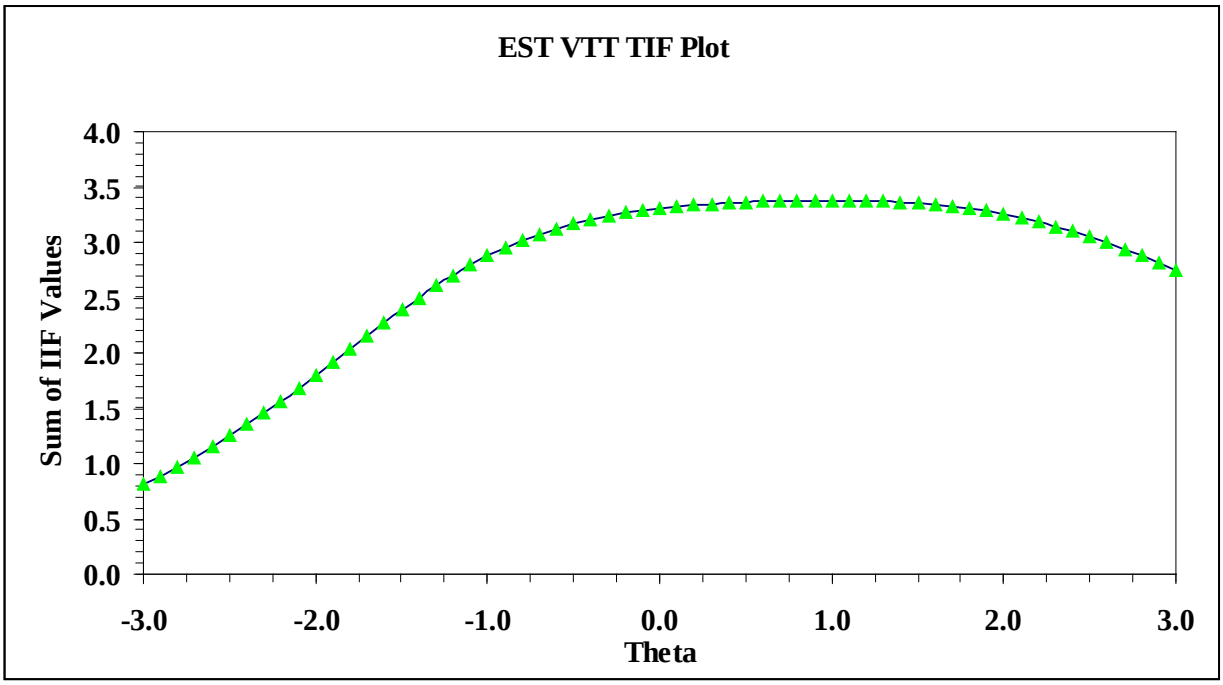


Figure 22. Test Information Function for the 9 VTT Items of the EST



Emergency Scenario CTT Item Level Analyses

Classical test theory analyses were performed on both dichotomous and constructed polytomous scores from the three emergency scenarios of the EST. As can be seen from Tables 60 and 61, both sets of scores produced similar results. The fire scenario was the easiest of the three, followed by the engine and propeller scenarios. Scores from all three scenarios were highly correlated indicating a single underlying ability. This is also evidenced by high factor loadings that resulted from fitting a single factor model to each set of scores (see the last column in each of the tables). These findings clearly suggest that the three scenario scores should be summed into a single total score to provide a more reliable index of examinee ability to respond to emergencies while under stress. Reliabilities of the resulting total scenario scores (EST Scenario Score and EST Scenario Poly) were 0.76 and 0.74, respectively. Interestingly, reliabilities, loadings and corrected item total correlations based on the original dichotomous scores tended to be higher than those based on constructed scores that weighted examinee scores by reaction time.

Table 60. CTT Statistics for the 3 Dichotomously Scored Emergency Scenarios of the EST

EST Item Name	Dichotomous Response			
	Mean	SD	CITC	Factor Loading
EP_EngineACC	.58	.49	.59	.71
EP_FireACC	.64	.48	.64	.81
EP_PropellerACC	.40	.49	.54	.63

Table 61. CTT Statistics for the 3 Polytomously Scored Emergency Scenarios of the EST

VTT Item Name	Polytomous Response			
	Mean	SD	CITC	Factor Loading
EngineRT_p	.89	.91	.60	.75
EP_FireRT_p	1.07	.95	.59	.74
EP_PropellerRT_p	.57	.79	.51	.60

EST Scale Scores

The subtest level indicators for the ATT, VTT, and emergency scenarios components of the EST were also examined separately for SPs, SNFOs, and the total sample. The total number of redirects (EST ATT Redirects and ES VTT Redirects), the average distance between the respective crosshairs and the targets during the test (EST ATT Average Distance and EST VTT Average Distance), the total numbers of on-target responses (EST ATT Total On Target and EST VTT Total On Target), the IRT score (EST ATT IRT Score and ESTT VTT IRT Score), and the two emergency scenario total scores (EST Scenario Total and EST Scenario Poly) are all indicators of examinee ability. As with the individually administered ATT and VTT assessments, the average distance measures are negatively related to the scores for the other components in the EST. The ATT and VTT components themselves correlated in the 0.40s and the correlations of both with emergency scenario component scores were generally in the 0.20s and 0.30s (see Table 65).

Tables 62 and 63 show descriptive statistics for the ATT and VTT components of the EST when analyzed separately using the total sample, SPs, and SNFOs. The SPs performed better on the ATT component by about a half of the total sample SD, but the effect size differences for VTT components were smaller. There were virtually no differences for emergency scenario scores, although SPs, when responding correctly, appeared to be a bit faster which is reflected in the higher mean of the EST Scenario Poly score for that group (see Table 64).

Table 62. Performance on the ATT Component of the EST Across SP and SNFO Student Groups

		EST ATT Redirects		EST ATT Average Distance		EST ATT Total On Target		EST ATT IRT Score	
Program	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SNFO	89	6.67	4.13	148.39	42.69	25.17	14.62	-0.23	0.88
SP	310	8.71	4.77	132.29	37.58	31.99	16.47	0.06	0.87
Total	399	8.25	4.71	135.88	39.30	30.47	16.30	0.00	0.88

Table 63. Performance on the VTT Component of the EST Across SP and SNFO Student Groups

		EST VTT Redirects		EST VTT Average Distance		EST VTT Total On Target		EST VTT IRT Score	
Program	N	Mean	SD	Mean	SD	Mean	SD	Mean	SD
SNFO	89	8.78	4.28	122.71	33.60	43.30	19.24	-0.12	0.86
SP	310	9.66	4.90	116.62	31.77	47.14	22.61	0.03	0.86
Total	399	9.46	4.78	117.98	32.24	46.29	21.94	0.00	0.86

Table 64. Performance on the Emergency Scenario Component of the EST Across SP and SNFO Student Groups

		EST Scenario Score		EST Scenario Poly	
Program	N	Mean	SD	Mean	SD
SNFO	89	1.62	1.23	2.36	2.05
SP	310	1.63	1.20	2.57	2.19
Total	399	1.62	1.20	2.53	2.16

Table 65. Correlations Between the Emergency Scenario, ATT, and VTT Component Scores of the EST

	EST ATT Redirects	EST ATT Average Distance	EST ATT Total On Target	EST ATT IRT Score	EST VTT Redirects	EST VTT Average Distance	EST VTT Total On Target	EST VTT IRT Score	EST Scenario Score	EST Scenario Poly
EST ATT Redirects	1	-.848**	.984**	.890**	.529**	-.527**	.526**	.444**	.227**	.259**
EST ATT Average Distance	-.848**	1	-.846**	-.810**	-.527**	.582**	-.522**	-.448**	-.288**	-.331**
EST ATT Total On Target	.984**	-.846**	1	.912**	.518**	-.520**	.517**	.446**	.226**	.258**
EST ATT IRT Score	.890**	-.810**	.912**	1	.450**	-.464**	.452**	.412**	.150**	.178**
EST VTT Redirects	.529**	-.527**	.518**	.450**	1	-.900**	.990**	.883**	.317**	.343**
EST VTT Average Distance	-.527**	.582**	-.520**	-.464**	-.900**	1	-.896**	-.846**	-.359**	-.391**
EST VTT Total On Target	.526**	-.522**	.517**	.452**	.990**	-.896**	1	.900**	.316**	.347**
EST VTT IRT Score	.444**	-.448**	.446**	.412**	.883**	-.846**	.900**	1	.202**	.233**
EST Scenario Score	.227**	-.288**	.226**	.150**	.317**	-.359**	.316**	.202**	1	.905**
EST Scenario Poly	.259**	-.331**	.258**	.178**	.343**	-.391**	.347**	.233**	.905**	1

Table 66 shows the correlations of the scores for the ATT components of the EST with other potential predictors of training performance. ATT component scores showed correlations in the 0.2 to 0.3 range with most criteria and a correlation of 0.35 with simExperience. As with the results for the ATTVTT and the MTT, the correlations with aTraining and Education were at or below 0.1. In general, these correlations mirrored the ATT results found in other PBM subtests.

Table 66. Correlations Between the ATT Component Scores of the EST and Other Predictors

	N	Mean	SD	EST ATT Redirects	EST ATT Average Distance	EST ATT Total On Target	EST ATT IRT Score
aTraining	390	0.23	0.72	0.098	-0.045	.107*	0.064
Education	385	2.88	0.59	.127*	-0.08	.133**	.108*
simExperience	399	0.74	0.83	.349**	-.255**	.348**	.276**
flightHours	391	0.69	1.38	0.088	-0.026	.102*	0.09
ANI_RAW	332	0.58	0.53	.134*	-.141*	.146**	.175**
MST_RAW	332	0.34	0.67	.168**	-.166**	.175**	.173**
RCT_RAW	332	0.43	0.53	0.052	-0.043	0.068	0.065
SAT_Post2004	332	0.76	0.64	.223**	-.229**	.216**	.189**
MCT_Post2004	332	0.50	0.64	.238**	-.255**	.247**	.240**
AQR_Post2004	332	0.55	0.52	.286**	-.296**	.299**	.302**
PFAR_Post2004	332	0.67	0.50	.273**	-.285**	.283**	.288**
FOFAR_Post2004	332	0.65	0.53	.279**	-.281**	.286**	.280**
OAR_Post2004	332	0.50	0.62	.247**	-.258**	.258**	.251**

Table 67 shows the correlations of the EST VTT component scores with other potential predictors of training performance. VTT component scores showed correlations that were generally in the 0.2 to 0.3 range with most criteria and a correlation of about 0.2 with simExperience. As with the results for the ATTVTT and the MTT, the correlations with aTraining and Education were at or below 0.1. In general, these correlations mirrored results for the VTT in other PBM subtests.

Table 67. Correlations Between the VTT Component Scores of the EST and Other Predictors

	N	Mean	SD	EST VTT Redirects	EST VTT Average Distance	EST VTT Total On Target	EST VTT IRT Score
aTraining	390	0.23	0.72	0.062	-0.037	0.069	0.053
Education	385	2.88	0.59	-0.028	-0.004	-0.044	-0.045
simExperience	399	0.74	0.83	.190**	-.201**	.193**	.169**
flightHours	391	0.69	1.38	-0.007	0.022	0.00	0.023
ANI_RAW	332	0.58	0.53	0.081	-.109*	0.084	.119*

MST_RAW	332	0.34	0.67	.215**	-.249**	.222**	.227**
RCT_RAW	332	0.43	0.53	.130*	-.152**	.147**	.150**
SAT_Post2004	332	0.76	0.64	.214**	-.257**	.201**	.165**
MCT_Post2004	332	0.50	0.64	.250**	-.301**	.243**	.178**
AQR_Post2004	332	0.55	0.52	.290**	-.350**	.290**	.272**
PFAR_Post2004	332	0.67	0.50	.246**	-.302**	.241**	.228**
FOFAR_Post2004	332	0.65	0.53	.294**	-.350**	.295**	.296**
OAR_Post2004	332	0.50	0.62	.280**	-.333**	.280**	.233**

Table 68 shows the correlations of the scenario component scores of the EST with other potential predictors of training performance. Most correlations were in the 0.1 to 0.2 range indicating that this component measured a different construct than the others.

Table 68. Correlations Between the Emergency Scenario Component Scores of the EST and Other Predictors

	N	Mean	Std. Deviation	EST Scenario Score	EST Scenario Poly
aTraining	390	0.23	0.72	.013	.002
Education	385	2.88	0.59	-.021	.024
simExperience	399	0.74	0.83	.037	.061
flightHours	391	0.69	1.38	-.043	-.064
ANI_RAW	332	0.58	0.53	.023	-.007
MST_RAW	332	0.34	0.67	.187**	.196**
RCT_RAW	332	0.43	0.53	.149**	.133*
SAT_Post2004	332	0.76	0.64	.088	.050
MCT_Post2004	332	0.50	0.64	.196**	.192**
AQR_Post2004	332	0.55	0.52	.204**	.185**
PFAR_Post2004	332	0.67	0.50	.140*	.107
FOFAR_Post2004	332	0.65	0.53	.188**	.161**
OAR_Post2004	332	0.50	0.62	.231**	.230**

Tables 69, 70, 71, 72, 73, and 74 present the correlations between the ATT, VTT, and emergency scenario component scores of the EST and training criteria (block grades and training composites) for the total sample and for student pilots only.

Similar to other subtests, the results for the SP sample shown in Table 70 indicated that the ATT component correlations were highest with Instruments_ALL, with values around .30; the corresponding correlations were smaller in the total sample. Note that the ATT components correlated around 0.25 with NSS.

The VTT results for SPs are shown in Table 72. The VTT Average Distance measures seemed to have somewhat higher correlations than the other measures. It correlated -0.33 with Instruments_ALL and -0.27 with NSS.

Overall, both the VTT and ATT components have correlations with criteria large enough to be of practical importance for selection. The emergency scenario correlations (Tables 73 and 74) were generally smaller than those for the ATT and VTT components (most were in the 0.10s). However, they may provide incremental validity for selection because of the correlations with Instruments and NSS grades in the 0.15 to 0.20 range and the relatively low correlations with the ATT and VTT component scores.

Table 69. Correlations Between the ATT Component Scores of the EST and Navy Pilot Training Criteria for the Total Sample

		Total Sample			
Training Block Name	N	EST ATT Redirects	EST ATT Average Distance	EST ATT Total On Target	EST ATT IRT Score
C20	399	.104*	-0.085	.112*	0.093
C40	86	0.031	-0.011	0.039	0.043
C41	374	.111*	-0.071	.113*	0.101
C42	367	0.089	-0.092	0.094	0.054
C43	270	0.064	-0.048	0.083	0.041
C45	262	.129*	-.132*	.144*	.132*
C46	246	.173**	-.191**	.162*	.145*
C47	238	0.116	-0.051	0.109	0.073
I20	387	.230**	-.219**	.232**	.217**
I21	300	.274**	-.279**	.271**	.247**
I22	207	.144*	-.187**	.165*	0.127
I23	206	.165*	-.186**	.184**	.139*
I24	194	.229**	-.225**	.239**	.220**
I25	188	0.139	-0.107	.157*	.151*
I40	380	.151**	-.181**	.155**	.152**
I41	282	.268**	-.269**	.286**	.272**
I42	233	.210**	-.173**	.210**	.186**
I43	227	0.09	-0.115	0.086	0.082
F40	225	.207**	-.211**	.213**	.183**
F42	209	.251**	-.252**	.250**	.251**
N40	183	0.133	-0.064	0.107	0.059
N41	181	0.023	-0.018	-0.001	-0.004
Contact_Simulation	399	.104*	-0.085	.112*	0.093

Contact_AIRCRAFT	378	.148**	-.125*	.154**	.127*
Contact_ALL	399	.137**	-.117*	.143**	.119*
Instruments_Simulation	387	.221**	-.215**	.228**	.193**
Instruments_AIRCRAFT	380	.219**	-.236**	.225**	.217**
Instruments_ALL	387	.242**	-.244**	.253**	.230**
Instruments_BASIC	387	.239**	-.239**	.241**	.226**
Instruments_RADIO	289	.240**	-.266**	.262**	.243**
Instruments_NAVIGATION	244	.212**	-.199**	.224**	.226**
Navigation_AIRCRAFT	183	0.095	-0.049	0.06	0.03
Formation_AIRCRAFT	225	.232**	-.249**	.239**	.224**
Navy Standard Score (NSS)	399	.202**	-.194**	.211**	.201**

Table 70. Correlations Between the ATT Component Scores of the EST and Navy Pilot Training Criteria for the Student Pilots Only

	Students Pilots (SPs)				
Training Block Name	N	EST ATT Redirects	EST ATT Average Distance	EST ATT Total On Target	EST ATT IRT Score
C20	310	.140*	-.124*	.147**	.121*
C40	-	-	-	-	-
C41	292	.170**	-.135*	.180**	.162**
C42	284	.128*	-.140*	.133*	0.097
C43	270	0.064	-0.048	0.083	0.041
C45	262	.129*	-.132*	.144*	.132*
C46	246	.173**	-.191**	.162*	.145*
C47	238	0.116	-0.051	0.109	0.073
I20	303	.273**	-.278**	.281**	.276**
I21	300	.274**	-.279**	.271**	.247**
I22	207	.144*	-.187**	.165*	0.127
I23	206	.165*	-.186**	.184**	.139*
I24	194	.229**	-.225**	.239**	.220**
I25	188	0.139	-0.107	.157*	.151*
I40	298	.189**	-.216**	.190**	.187**
I41	200	.237**	-.230**	.256**	.223**
I42	183	.248**	-.237**	.248**	.210**
I43	178	0.049	-0.066	0.04	0.004
F40	225	.207**	-.211**	.213**	.183**
F42	209	.251**	-.252**	.250**	.251**

N40	183	0.133	-0.064	0.107	0.059
N41	181	0.023	-0.018	-0.001	-0.004
Contact_Simulation	310	.140*	-.124*	.147**	.121*
Contact_AIRCRAFT	292	.195**	-.183**	.202**	.171**
Contact_ALL	310	.176**	-.165**	.183**	.153**
Instruments_Simulation	303	.269**	-.280**	.284**	.255**
Instruments_AIRCRAFT	298	.271**	-.294**	.275**	.259**
Instruments_ALL	303	.293**	-.306**	.306**	.281**
Instruments_BASIC	303	.291**	-.300**	.296**	.287**
Instruments_RADIO	207	.204**	-.229**	.227**	.183**
Instruments_NAVIGATION	194	.222**	-.222**	.235**	.220**
Navigation_AIRCRAFT	183	0.095	-0.049	0.06	0.03
Formation_AIRCRAFT	225	.232**	-.249**	.239**	.224**
Navy Standard Score (NSS)	310	.240**	-.243**	.250**	.232**

Table 71. Correlations Between the VTT Component Scores of the EST and Navy Pilot Training Criteria for the Total Sample

	Total Sample				
Training Block Name	N	EST VTT Redirects	EST VTT Average Distance	EST VTT Total On Target	EST VTT IRT Score
C20	399	0.038	-0.062	0.042	0.02
C40	86	0.178	-0.079	0.182	0.118
C41	374	0.086	-0.099	0.09	0.077
C42	367	0.077	-0.096	0.085	0.072
C43	270	0.008	-0.047	0.019	0.05
C45	262	0.101	-0.086	0.113	0.121
C46	246	0.074	-0.12	0.073	0.105
C47	238	0.079	-0.106	0.072	0.117
I20	387	.255**	-.309**	.255**	.247**
I21	300	.269**	-.336**	.274**	.235**
I22	207	0.086	-.169*	0.099	0.102
I23	206	.157*	-.222**	.181**	.173*
I24	194	.168*	-.190**	.179*	.194**
I25	188	0.04	-0.075	0.053	0.099
I40	380	.223**	-.223**	.218**	.197**
I41	282	.202**	-.232**	.215**	.181**
I42	233	0.111	-.143*	0.125	.152*

I43	227	0.053	-0.104	0.072	0.094
F40	225	.221**	-.266**	.206**	.170*
F42	209	.211**	-.253**	.206**	.180**
N40	183	0.056	-0.048	0.045	0.04
N41	181	0.022	-0.011	0.014	0.056
Contact_Simulation	399	0.038	-0.062	0.042	0.02
Contact_AIRCRAFT	378	.136**	-.138**	.140**	.136**
Contact_ALL	399	.099*	-.116*	.108*	0.091
Instruments_Simulation	387	.230**	-.284**	.234**	.231**
Instruments_AIRCRAFT	380	.208**	-.226**	.213**	.184**
Instruments_ALL	387	.245**	-.295**	.253**	.247**
Instruments_BASIC	387	.294**	-.331**	.296**	.278**
Instruments_RADIO	289	.185**	-.239**	.202**	.180**
Instruments_NAVIGATION	244	0.123	-.175**	.141*	.177**
Navigation_AIRCRAFT	183	0.05	-0.046	0.042	0.068
Formation_AIRCRAFT	225	.237**	-.288**	.221**	.184**
Navy Standard Score (NSS)	399	.192**	-.233**	.201**	.182**

Table 72. Correlations Between the VTT Component Scores of the EST and Navy Pilot Training Criteria for the Student Pilots Only

	Students Pilots (SPs)				
Training Block Name	N	EST VTT Redirects	EST VTT Average Distance	EST VTT Total On Target	EST VTT IRT Score
C20	310	0.049	-0.087	0.05	0.015
C40	-	-	-	-	-
C41	292	.152**	-.173**	.156**	.148*
C42	284	.122*	-.166**	.129*	.143*
C43	270	0.008	-0.047	0.019	0.05
C45	262	0.101	-0.086	0.113	0.121
C46	246	0.074	-0.12	0.073	0.105
C47	238	0.079	-0.106	0.072	0.117
I20	303	.262**	-.344**	.266**	.268**
I21	300	.269**	-.336**	.274**	.235**
I22	207	0.086	-.169*	0.099	0.102
I23	206	.157*	-.222**	.181**	.173*
I24	194	.168*	-.190**	.179*	.194**
I25	188	0.04	-0.075	0.053	0.099

I40	298	.245**	-.251**	.239**	.219**
I41	200	.179*	-.201**	.194**	.160*
I42	183	0.133	-.159*	0.142	.146*
I43	178	0.03	-0.063	0.047	0.076
F40	225	.221**	-.266**	.206**	.170*
F42	209	.211**	-.253**	.206**	.180**
N40	183	0.056	-0.048	0.045	0.04
N41	181	0.022	-0.011	0.014	0.056
Contact_Simulation	310	0.049	-0.087	0.05	0.015
Contact_AIRCRAFT	292	.144*	-.182**	.148*	.169**
Contact_ALL	310	0.103	-.146**	.113*	0.103
Instruments_Simulation	303	.233**	-.316**	.243**	.250**
Instruments_AIRCRAFT	298	.230**	-.253**	.234**	.197**
Instruments_ALL	303	.258**	-.328**	.268**	.266**
Instruments_BASIC	303	.315**	-.378**	.319**	.307**
Instruments_RADIO	207	.157*	-.213**	.177*	.160*
Instruments_NAVIGATION	194	0.131	-.172*	.144*	.173*
Navigation_AIRCRAFT	183	0.05	-0.046	0.042	0.068
Formation_AIRCRAFT	225	.237**	-.288**	.221**	.184**
Navy Standard Score (NSS)	310	.209**	-.273**	.219**	.200**

Table 73. Correlations Between the Emergency Scenario Component Scores of the EST and Navy Pilot Training Criteria for the Total Sample

	Total Sample		
Training Block Name	N	EST Scenario Score	EST Scenario Poly
C20	399	.127*	.122*
C40	86	0.184	0.129
C41	374	0.054	0.038
C42	367	0.063	0.067
C43	270	-0.027	0.002
C45	262	0.05	0.07
C46	246	-0.001	0.02
C47	238	-0.041	-0.039
I20	387	.146**	.130*
I21	300	.159**	.168**
I22	207	0.118	.144*
I23	206	.186**	.168*

I24	194	0.095	0.11
I25	188	0.032	0.038
I40	380	.128*	.122*
I41	282	.187**	.137*
I42	233	0.122	0.07
I43	227	0.11	0.065
F40	225	0.076	0.112
F42	209	0.06	0.083
N40	183	0.027	-0.002
N41	181	0.055	0.01
Contact_Simulation	399	.127*	.122*
Contact_AIRCRAFT	378	0.088	0.073
Contact_ALL	399	.117*	.116*
Instruments_Simulation	387	.173**	.157**
Instruments_AIRCRAFT	380	.166**	.125*
Instruments_ALL	387	.192**	.165**
Instruments_BASIC	387	.176**	.166**
Instruments_RADIO	289	.190**	.149*
Instruments_NAVIGATION	244	0.114	0.072
Navigation_AIRCRAFT	183	0.057	0.003
Formation_AIRCRAFT	225	0.08	0.107
Navy Standard Score (NSS)	399	.179**	.158**

Table 74. Correlations Between the Emergency Scenario Component Scores of the EST and Navy Pilot Training Criteria for the Student Pilots Only

	Students Pilots (SPs)		
Training Block Name	N	EST Scenario Score	EST Scenario Poly
C20	310	.117*	.121*
C40	0	-	-
C41	292	0.073	0.069
C42	284	0.035	0.049
C43	270	-0.027	0.002
C45	262	0.05	0.07
C46	246	-0.001	0.02
C47	238	-0.041	-0.039
I20	303	.131*	.132*
I21	300	.159**	.168**

I22	207	0.118	.144*
I23	206	.186**	.168*
I24	194	0.095	0.11
I25	188	0.032	0.038
I40	298	0.099	0.099
I41	200	.165*	.141*
I42	183	0.129	0.094
I43	178	0.067	0.038
F40	225	0.076	0.112
F42	209	0.06	0.083
N40	183	0.027	-0.002
N41	181	0.055	0.01
Contact_Simulation	310	.117*	.121*
Contact_AIRCRAFT	292	0.05	0.053
Contact_ALL	310	0.089	0.109
Instruments_Simulation	303	.163**	.167**
Instruments_AIRCRAFT	298	.151**	.122*
Instruments_ALL	303	.176**	.167**
Instruments_BASIC	303	.158**	.162**
Instruments_RADIO	207	.170*	.161*
Instruments_NAVIGATION	194	0.09	0.07
Navigation_AIRCRAFT	183	0.057	0.003
Formation_AIRCRAFT	225	0.08	0.107
Navy Standard Score (NSS)	310	.156**	.160**

Table 75 presents the results of multiple regression analyses using the ATT, VTT, and emergency scenario components of the EST as predictors of composite performance criteria for the student pilot sample. Prior to these analyses, we checked for the interaction between ATT and VTT components by conducting a series of moderated regression analyses (not shown here). In all 20 analyses conducted, no significant interactions were found, so we did not include these interaction terms in analyses described in Table 75.

Similar to the ATTVTT and MTT, the ATT component scores of the EST significantly predicted four of the five training performance criteria (all except Navigation grades). The VTT component was also significant in several analyses involving Instruments, Formation, and NSS grades. The Emergency Scenario component was significant in only two of the 20 analyses. In these two analyses, the standardized beta coefficients were reasonably large, so this subtest might be considered for inclusion in a future selection composite.

Table 75. EST Multiple Regression Results using the Emergency Scenario, ATT, and VTT Component Scores as Predictors of Navy Pilot Training Criteria

	Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	R
		B	Std. Error	Beta			
Contact ALL	(Constant)	46.50	1.03		45.20	0.00	0.183
	EST ATT Redirects	0.25	0.10	0.16	2.46	0.01	
	EST VTT Redirects	0.00	0.10	0.00	0.02	0.99	
	EST Scenario Score	0.31	0.36	0.05	0.85	0.40	
	(Constant)	53.82	2.15		25.03	0.00	0.179
	EST ATT Average Distance	-0.02	0.02	-0.07	-1.01	0.31	
	EST VTT Average Distance	-0.02	0.01	-0.12	-1.70	0.09	
	EST Scenario Score	0.18	0.37	0.03	0.48	0.63	
	(Constant)	46.17	1.07		42.98	0.00	0.190
	EST ATT Total On Target	0.07	0.03	0.17	2.51	0.01	
	EST VTT Total On Target	0.00	0.02	0.01	0.21	0.83	
	EST Scenario Score	0.28	0.36	0.05	0.78	0.43	
	(Constant)	48.51	0.70		69.07	0.00	0.172
	EST ATT IRT Score	1.07	0.51	0.13	2.09	0.04	
	EST VTT IRT Score	0.37	0.53	0.04	0.70	0.48	
	EST Scenario Score	0.38	0.35	0.06	1.08	0.28	
Instruments ALL	(Constant)	42.84	1.13		38.07	0.00	0.329
	EST ATT Redirects	0.36	0.11	0.21	3.25	0.00	
	EST VTT Redirects	0.20	0.11	0.12	1.80	0.07	
	EST Scenario Score	0.64	0.40	0.09	1.63	0.11	
	(Constant)	59.81	2.30		25.95	0.00	0.363
	EST ATT Average Distance	-0.06	0.02	-0.21	-3.20	0.00	
	EST VTT Average Distance	-0.04	0.01	-0.17	-2.63	0.01	
	EST Scenario Score	0.36	0.40	0.05	0.89	0.37	
	(Constant)	42.23	1.17		36.08	0.00	0.343
	EST ATT Total On Target	0.11	0.03	0.22	3.50	0.00	
	EST VTT Total On Target	0.05	0.02	0.13	1.95	0.05	
	EST Scenario Score	0.62	0.39	0.09	1.56	0.12	
	(Constant)	47.46	0.76		62.48	0.00	0.350
	EST ATT IRT Score	1.89	0.55	0.20	3.41	0.00	

	EST VTT IRT Score	1.63	0.57	0.17	2.86	0.00	
	EST Scenario Score	0.82	0.38	0.12	2.16	0.03	
Navigation AIRCRAFT	(Constant)	48.56	1.65		29.34	0.00	0.104
	EST ATT Redirects	0.16	0.15	0.09	1.07	0.29	
	EST VTT Redirects	-0.02	0.16	-0.01	-0.10	0.92	
	EST Scenario Score	0.32	0.56	0.04	0.58	0.56	
	(Constant)	51.30	3.31		15.50	0.00	0.070
	EST ATT Average Distance	0.00	0.02	-0.01	-0.17	0.87	
	EST VTT Average Distance	-0.01	0.02	-0.03	-0.38	0.71	
	EST Scenario Score	0.33	0.57	0.05	0.58	0.56	
	(Constant)	48.89	1.74		28.13	0.00	0.077
	EST ATT Total On Target	0.03	0.04	0.05	0.60	0.55	
	EST VTT Total On Target	0.00	0.03	0.00	0.03	0.98	
	EST Scenario Score	0.35	0.56	0.05	0.62	0.54	
	(Constant)	49.74	1.10		45.18	0.00	0.082
	EST ATT IRT Score	0.04	0.76	0.00	0.05	0.96	
	EST VTT IRT Score	0.60	0.84	0.06	0.72	0.47	
	EST Scenario Score	0.34	0.54	0.05	0.62	0.54	
Formation AIRCRAFT	(Constant)	44.73	1.48		30.19	0.00	0.271
	EST ATT Redirects	0.29	0.14	0.15	2.04	0.04	
	EST VTT Redirects	0.30	0.15	0.16	2.09	0.04	
	EST Scenario Score	-0.02	0.51	0.00	-0.03	0.97	
	(Constant)	62.58	2.99		20.95	0.00	0.312
	EST ATT Average Distance	-0.07	0.02	-0.23	-2.93	0.00	
	EST VTT Average Distance	-0.03	0.02	-0.14	-1.82	0.07	
	EST Scenario Score	-0.29	0.51	-0.04	-0.57	0.57	
	(Constant)	44.41	1.56		28.43	0.00	0.268
	EST ATT Total On Target	0.09	0.04	0.17	2.33	0.02	
	EST VTT Total On Target	0.06	0.03	0.14	1.80	0.07	
	EST Scenario Score	0.02	0.51	0.00	0.04	0.97	
	(Constant)	49.48	0.97		50.94	0.00	0.253
	EST ATT IRT Score	1.78	0.69	0.18	2.56	0.01	
	EST VTT IRT Score	1.23	0.77	0.11	1.61	0.11	
	EST Scenario Score	0.30	0.49	0.04	0.61	0.54	
Navy Standard	(Constant)	43.30	1.33		32.52	0.00	0.272

Score (NSS)	EST ATT Redirects	0.35	0.13	0.17	2.64	0.01	
	EST VTT Redirects	0.18	0.13	0.09	1.37	0.17	
	EST Scenario Score	0.70	0.47	0.09	1.49	0.14	
	(Constant)	59.33	2.76		21.50	0.00	0.298
	EST ATT Average Distance	-0.06	0.02	-0.18	-2.72	0.01	
	EST VTT Average Distance	-0.03	0.02	-0.12	-1.86	0.06	
	EST Scenario Score	0.43	0.48	0.05	0.90	0.37	
	(Constant)	42.72	1.39		30.78	0.00	0.282
	EST ATT Total On Target	0.11	0.04	0.18	2.79	0.01	
	EST VTT Total On Target	0.04	0.03	0.10	1.54	0.13	
	EST Scenario Score	0.67	0.47	0.08	1.43	0.15	
	(Constant)	47.64	0.90		52.68	0.00	0.283
	EST ATT IRT Score	1.91	0.66	0.17	2.89	0.00	
	EST VTT IRT Score	1.31	0.68	0.12	1.93	0.05	
	EST Scenario Score	0.88	0.45	0.11	1.93	0.05	

In summary, the EST provides valuable information for decision making. Similar to other subtests involving tracking tasks, both the ATT and VTT components showed reasonable discrimination parameters and a wide range of difficulties when analyzed using CTT and IRT methods. Both showed moderate criterion correlations and appeared to contribute independently to the prediction of training grades as was evident from insignificant interaction terms in the moderated multiple regression analyses. The patterns of validities in this and other tracking subtests were very similar, suggesting that aggregating ATT and VTT scores across tasks might improve reliability and validity. Alternatively, some of the tasks might be deleted if administration time is a concern.

Emergency Scenario scores, although based on just three events, showed a modest potential for predicting training grades. Because Emergency scenario scores did not correlate highly with other predictors, it may be useful to include this component into a future selection battery.

COMBINED ATT AND VTT SCORES AND THEIR VALIDITIES

Analyses of the VTT, ATT, ATTVTT, MTT, and EST subtests indicated that scores on one-dimensional (VTT) and two-dimensional (ATT) tracking tasks related to training criteria in similar ways. Moreover, although the difficulty of each subtest progressively increased, the rank order of examinees within each tracking task remained relatively unchanged. Both of these findings indicate that psychomotor abilities measured by these subtests are largely the same, so for selection purposes, it might make sense to aggregate ATT and VTT components into larger, more reliable composites. By doing so, one would not only reduce the number of potential ATT and VTT predictors, but also likely increase criterion validities of the respective components.

The aggregate analyses, presented below, are relatively straightforward. At the PBM level, each ATT and VTT subtest score was treated as an "item", so CTT analyses could be conducted to see if aggregation was justified. Next, test-level composites were created by summing (or averaging) across the respective subtest scores. Because scores on different subtests had different standard deviations, they were all standardized prior to forming composites. Finally, criterion related validities and regression analyses were conducted in a manner similar to those for individual subtests.

Item Level Analyses of the ATT and VTT Subtest Scores

Table 76 presents CTT statistics for four kinds of ATT scores (Redirects, Average Distance, on Target, and IRT) found in the ATT, ATTVTT, MTT and EST subtests. As can be seen, combining the subtest scores resulted in highly homogeneous composites: Each four "item" measure has a reliability of 0.92 or greater. Table 76 shows very high CITCs and factor loadings as well. Interestingly, the best indicators of two-dimensional tracking appear to result from the ATTVTT and MTT.

Table 76. CTT Statistics for the ATT Component Scores from the ATT, ATTVTT, MTT, and EST

ATT Subtest Name	Mean	SD	CITC	Factor Loading	Alpha
ATT Redirects	8.80	4.31	.83	.85	0.93
ATTVTT ATT Redirects	9.37	5.50	.92	.95	
MTT ATT Redirects	15.62	8.23	.92	.96	
EST ATT Redirects	8.24	4.71	.86	.87	
ATT Average Distance	71.70	27.10	.79	.82	0.93
ATTVTT ATT Average Distance	114.65	35.65	.90	.95	
MTT ATT Average Distance	107.40	37.08	.90	.95	
EST ATT Average Distance	135.92	39.35	.77	.80	
ATT Total On Target	32.22	14.68	.81	.83	0.92
ATTVTT ATT Total On	34.55	19.17	.91	.94	

Target					
MTT ATT Total On Target	55.94	28.30	.91	.95	
EST ATT Total On Target	30.44	16.32	.84	.86	
ATT IRT Score	.00	.92	.78	.82	0.92
ATTVTT ATT IRT Score	.00	.90	.84	.90	
MTT ATT IRT Score	.00	.91	.85	.90	
EST ATT IRT Score	.00	.88	.79	.83	

Table 77 presents VTT statistics for the four kinds of VTT scores (Redirects, Average Distance, on Target, and IRT) obtained from the VTT, ATTVTT, MTT and EST. In comparison to ATT scores, the VTT scores were somewhat less homogeneous, with reliabilities in the mid 0.80s. The corrected item-total correlations were high, but lower than the corresponding ATT values. The VTT scores obtained from the MTT had the highest corrected item-total correlation and factor loading in all analyses. In sum, all the CTT analyses indicate that there is a compelling case for forming aggregate ATT and VTT scores.

Table 77. CTT Statistics for the VTT Component Scores from ATT, ATTVTT, MTT, and EST Subtests

VTT Subtest Name	Mean	SD	CITC	Factor Loading	Alpha
VTT Redirects	17.15	3.25	.54	.57	0.85
ATTVTT VTT Redirects	12.51	4.51	.77	.83	
MTT VTT Redirects	16.54	7.04	.83	.92	
EST VTT Redirects	9.43	4.77	.76	.81	
VTT Average Distance	25.71	9.94	.50	.53	0.85
ATTVTT VTT Average Distance	87.66	24.17	.81	.88	
MTT VTT Average Distance	100.03	30.17	.85	.91	
EST VTT Average Distance	118.15	32.24	.79	.84	
VTT Total On Target	81.29	14.95	.51	.54	0.84
ATTVTT VTT Total On Target	60.46	20.85	.76	.81	
MTT VTT Total On Target	79.10	32.52	.82	.92	
EST VTT Total On Target	46.15	21.89	.76	.81	
VTT IRT Score	.01	.90	.50	.55	0.83
ATTVTT VTT IRT Score	.00	.84	.68	.77	
MTT VTT IRT Score	.00	.88	.73	.85	
EST VTT IRT Score	.00	.86	.70	.79	

Table 78 shows intercorrelations between eight resulting standardized ATT and VTT composites (four composites for ATT and four composites for VTT). As can be seen, all ATT composites correlated 0.9 and above with ATT Redirects and ATT on Target composites correlating 0.996.

The VTT composites were similarly intercorrelated. The ATT composites correlated in the 0.5s and 0.6s with the VTT composites.

Table 78. Intercorrelations Among the ATT and VTT Composites

	ATT Redirects Z	VTT Redirects Z	ATT Average Distance Z	VTT Average Distance Z	ATT On Target Z	VTT On Target Z	ATT IRT Z	VTT IRT Z
ATT Redirects Z	1	.655**	-.898**	-.642**	.996**	.652**	.965**	.619**
VTT Redirects Z	.655**	1	-.570**	-.940**	.657**	.996**	.609**	.962**
ATT Average Distance Z	-.898**	-.570**	1	.621**	-.901**	-.566**	-.926**	-.554**
VTT Average Distance Z	-.642**	-.940**	.621**	1	-.645**	-.937**	-.618**	-.938**
ATT On Target Z	.996**	.657**	-.901**	-.645**	1	.654**	.972**	.623**
VTT On Target Z	.652**	.996**	-.566**	-.937**	.654**	1	.605**	.966**
ATT IRT Z	.965**	.609**	-.926**	-.618**	.972**	.605**	1	.590**
VTT IRT Z	.619**	.962**	-.554**	-.938**	.623**	.966**	.590**	1

Table 79 shows correlations of other predictors with the ATT and VTT composites. The general level of the correlations in this table is somewhat higher than the correlations seen in corresponding tables for each of the separate tasks. This effect is probably due to the increased reliability of the composites. Note that both ATT Redirects and ATT On Target composites tend to correlate higher with simExperience than the Average Distance and IRT Composites.

Table 79. Correlations Between the ATT and VTT Composites and Other Predictors

	N	Mean	SD	ATT Redirects Z	VTT Redirects Z	ATT Average Distance Z	VTT Average Distance Z	ATT On Target Z	VTT On Target Z	ATT IRT Z	VTT IRT Z
aTraining	390	0.23	0.72	0.095	0.053	-0.073	-0.047	0.097	0.055	0.086	0.036
Education	385	2.88	0.59	0.094	0.019	-0.068	-0.027	0.097	0.016	0.088	0.026
simExperience	399	0.74	0.83	.423**	.294**	-.351**	-.270**	.423**	.296**	.388**	.260**
flightHours	391	0.69	1.38	.137**	0.047	-0.088	-0.017	.143**	0.048	.126*	0.036
ANI_RAW	332	0.58	0.53	.215**	.146**	-.218**	-.181**	.218**	.148**	.220**	.160**
MST_RAW	332	0.34	0.67	.142**	.214**	-.127*	-.219**	.150**	.213**	.148**	.208**
RCT_RAW	332	0.43	0.53	0.056	.128*	-0.054	-.141*	0.058	.136*	0.053	.140*
SAT_Post2004	332	0.76	0.64	.238**	.239**	-.233**	-.283**	.243**	.240**	.235**	.234**
MCT_Post2004	332	0.50	0.64	.240**	.250**	-.233**	-.265**	.248**	.246**	.235**	.226**
AQR_Post2004	332	0.55	0.52	.312**	.320**	-.304**	-.353**	.322**	.320**	.313**	.315**
PFAR_Post2004	332	0.67	0.50	.328**	.296**	-.325**	-.341**	.336**	.296**	.329**	.295**
FOFAR_Post2004	332	0.65	0.53	.296**	.327**	-.287**	-.368**	.306**	.330**	.299**	.330**
OAR_Post2004	332	0.50	0.62	.238**	.280**	-.228**	-.294**	.248**	.278**	.237**	.261**

Tables 80, 81, 82, and 83 present criterion-related validities for the ATT and VTT composites formed from the total sample and with student pilots only. As can be seen, criterion validities are generally similar to those that were observed for the individual PBM subtests. Thus, the increased reliabilities of the composites did not appear to boost validity coefficients substantially, suggesting that a shortened test battery might be as effective as the full length assessment.

Table 80. Correlations Between the ATT Composite Scores and Navy Pilot Training Criteria for the Total Sample

		Total Sample			
Training Block Name	N	ATT Redirects Z	ATT Average Distance Z	ATT OnTarget Z	ATT IRT Z
C20	399	0.089	-0.077	0.095	0.073
C40	86	0.047	-0.011	0.047	0.04
C41	374	.124*	-0.068	.126*	.111*
C42	367	0.085	-0.079	0.091	0.077
C43	270	0.094	-0.094	0.1	0.083
C45	262	.158*	-.168**	.167**	.173**
C46	246	.176**	-.214**	.175**	.158*
C47	238	.152*	-0.102	.154*	0.126
I20	387	.279**	-.252**	.284**	.270**
I21	300	.274**	-.277**	.286**	.266**
I22	207	.138*	-.185**	.147*	.137*
I23	206	.162*	-.178*	.167*	.154*
I24	194	.237**	-.230**	.242**	.229**
I25	188	.170*	-.165*	.175*	.173*
I40	380	.145**	-.153**	.149**	.137**
I41	282	.256**	-.265**	.266**	.257**
I42	233	.210**	-.181**	.215**	.198**
I43	227	0.106	-0.11	0.107	0.099
F40	225	.225**	-.236**	.233**	.202**
F42	209	.268**	-.275**	.268**	.260**
N40	183	0.101	-0.047	0.099	0.074
N41	181	-0.011	0.013	-0.01	-0.007
Contact_Simulation	399	0.089	-0.077	0.095	0.073
Contact_AIRCRAFT	378	.165**	-.130*	.168**	.149**
Contact_ALL	399	.148**	-.124*	.154**	.136**
Instruments_Simulation	387	.248**	-.226**	.255**	.238**
Instruments_AIRCRAFT	380	.204**	-.201**	.208**	.189**

Instruments_ALL	387	.254**	-.243**	.262**	.244**
Instruments_BASIC	387	.262**	-.245**	.269**	.253**
Instruments_RADIO	289	.219**	-.251**	.230**	.223**
Instruments_NAVIGATION	244	.227**	-.222**	.233**	.230**
Navigation_AIRCRAFT	183	0.055	-0.018	0.051	0.039
Formation_AIRCRAFT	225	.247**	-.268**	.255**	.233**
Navy Standard Score (NSS)	399	.226**	-.217**	.235**	.218**

Table 81. Correlations Between the ATT Composite Scores and Navy Pilot Training Criteria for the Student Pilots Only

	Students Pilots (SPs)				
Training Block Name	N	ATT Redirects Z	ATT Average Distance Z	ATT OnTarget Z	ATT IRT Z
C20	310	.118*	-.118*	.124*	0.104
C40	-	-	-	-	-
C41	292	.183**	-.142*	.188**	.168**
C42	284	.146*	-.135*	.147*	.133*
C43	270	0.094	-0.094	0.1	0.083
C45	262	.158*	-.168**	.167**	.173**
C46	246	.176**	-.214**	.175**	.158*
C47	238	.152*	-0.102	.154*	0.126
I20	303	.322**	-.324**	.328**	.319**
I21	300	.274**	-.277**	.286**	.266**
I22	207	.138*	-.185**	.147*	.137*
I23	206	.162*	-.178*	.167*	.154*
I24	194	.237**	-.230**	.242**	.229**
I25	188	.170*	-.165*	.175*	.173*
I40	298	.179**	-.190**	.181**	.171**
I41	200	.246**	-.243**	.254**	.240**
I42	183	.236**	-.213**	.239**	.218**
I43	178	0.079	-0.078	0.076	0.058
F40	225	.225**	-.236**	.233**	.202**
F42	209	.268**	-.275**	.268**	.260**
N40	183	0.101	-0.047	0.099	0.074
N41	181	-0.011	0.013	-0.01	-0.007
Contact_Simulation	310	.118*	-.118*	.124*	0.104
Contact_AIRCRAFT	292	.223**	-.200**	.226**	.208**

Contact_ALL	310	.188**	-.179**	.195**	.178**
Instruments_Simulation	303	.287**	-.296**	.297**	.282**
Instruments_AIRCRAFT	298	.249**	-.255**	.252**	.233**
Instruments_ALL	303	.299**	-.308**	.308**	.293**
Instruments_BASIC	303	.308**	-.315**	.317**	.305**
Instruments_RADIO	207	.199**	-.228**	.208**	.195**
Instruments_NAVIGATION	194	.239**	-.240**	.244**	.237**
Navigation_AIRCRAFT	183	0.055	-0.018	0.051	0.039
Formation_AIRCRAFT	225	.247**	-.268**	.255**	.233**
Navy Standard Score (NSS)	310	.269**	-.277**	.277**	.260**

Table 82. Correlations Between the VTT Composite Scores and Navy Pilot Training Criteria for the Total Sample

	Total Sample				
Training Block Name	N	VTT Redirects Z	VTT Average Distance Z	VTT OnTarget Z	VTT IRT Z
C20	399	0.061	-0.067	0.061	0.052
C40	86	0.175	-0.145	0.169	0.166
C41	374	0.098	-0.101	0.101	0.079
C42	367	0.095	-.107*	0.098	0.089
C43	270	0.067	-0.077	0.064	0.065
C45	262	.138*	-0.12	.141*	.142*
C46	246	0.102	-0.122	0.096	0.096
C47	238	.149*	-.132*	.135*	.139*
I20	387	.280**	-.282**	.275**	.258**
I21	300	.294**	-.310**	.290**	.272**
I22	207	0.103	-.145*	0.109	0.094
I23	206	0.126	-.162*	0.129	0.093
I24	194	.213**	-.207**	.209**	.193**
I25	188	0.092	-0.115	0.089	0.095
I40	380	.210**	-.199**	.211**	.197**
I41	282	.184**	-.221**	.188**	.162**
I42	233	.139*	-.145*	.139*	.143*
I43	227	0.072	-0.101	0.073	0.076
F40	225	.252**	-.270**	.257**	.243**
F42	209	.256**	-.281**	.258**	.254**
N40	183	0.108	-0.108	0.11	0.115

N41	181	0.064	-0.063	0.068	0.08
Contact_Simulation	399	0.061	-0.067	0.061	0.052
Contact_AIRCRAFT	378	.169**	-.169**	.168**	.162**
Contact_ALL	399	.131**	-.134**	.131**	.123*
Instruments_Simulation	387	.255**	-.262**	.252**	.230**
Instruments_AIRCRAFT	380	.190**	-.190**	.191**	.168**
Instruments_ALL	387	.252**	-.262**	.252**	.231**
Instruments_BASIC	387	.302**	-.297**	.299**	.279**
Instruments_RADIO	289	.156**	-.209**	.162**	.136*
Instruments_NAVIGATION	244	.165*	-.183**	.160*	.155*
Navigation_AIRCRAFT	183	0.106	-0.107	0.108	0.117
Formation_AIRCRAFT	225	.266**	-.296**	.270**	.260**
Navy Standard Score (NSS)	399	.218**	-.230**	.219**	.198**

Table 83. Correlations between the VTT Composite Scores and Navy Pilot Training Criteria for the Students Sample Only

		Students Pilots (SPs)			
Training Block Name	N	VTT Redirects Z	VTT Average Distance Z	VTT OnTarget Z	VTT IRT Z
C20	310	0.075	-0.082	0.078	0.059
C40	-	-	-	-	-
C41	292	.172**	-.168**	.175**	.156**
C42	284	.165**	-.189**	.169**	.163**
C43	270	0.067	-0.077	0.064	0.065
C45	262	.138*	-0.12	.141*	.142*
C46	246	0.102	-0.122	0.096	0.096
C47	238	.149*	-.132*	.135*	.139*
I20	303	.297**	-.317**	.292**	.285**
I21	300	.294**	-.310**	.290**	.272**
I22	207	0.103	-.145*	0.109	0.094
I23	206	0.126	-.162*	0.129	0.093
I24	194	.213**	-.207**	.209**	.193**
I25	188	0.092	-0.115	0.089	0.095
I40	298	.240**	-.226**	.240**	.226**
I41	200	.186**	-.197**	.190**	.161*
I42	183	.150*	-.155*	.149*	.148*
I43	178	0.065	-0.075	0.064	0.07

F40	225	.252**	-.270**	.257**	.243**
F42	209	.256**	-.281**	.258**	.254**
N40	183	0.108	-0.108	0.11	0.115
N41	181	0.064	-0.063	0.068	0.08
Contact_Simulation	310	0.075	-0.082	0.078	0.059
Contact_AIRCRAFT	292	.196**	-.206**	.195**	.191**
Contact_ALL	310	.143*	-.151**	.146*	.135*
Instruments_Simulation	303	.266**	-.294**	.264**	.251**
Instruments_AIRCRAFT	298	.222**	-.218**	.224**	.196**
Instruments_ALL	303	.277**	-.297**	.278**	.260**
Instruments_BASIC	303	.332**	-.341**	.330**	.316**
Instruments_RADIO	207	.150*	-.182**	.157*	0.127
Instruments_NAVIGATION	194	.179*	-.188**	.172*	.164*
Navigation_AIRCRAFT	183	0.106	-0.107	0.108	0.117
Formation_AIRCRAFT	225	.266**	-.296**	.270**	.260**
Navy Standard Score (NSS)	310	.248**	-.264**	.250**	.226**

Finally, in Table 84, we present regression results for using ATT and VTT composite scores in predicting the five training grade composites for the SP sample only. With the exception of Navigation, both ATT and VTT composites appear to contribute to predictive power. ATT appears more important for predicting Contact ALL, Instruments ALL, and NSS, while VTT appears more important for Formation ALL. Of the four approaches to scoring, Average Distance composites generally had the highest validity, but the differences were so small as to make any of the approaches viable. The choice should probably be guided by the "ease of use" and "computational" considerations.

Table 84. Multiple Regression Results for Four Types of Standardized ATT and VTT Composite Scores as Predictors of Navy Pilot Training Criteria

	Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	R
		B	Std. Error	Beta			
Contact ALL	(Constant)	49.14	0.41		119.37	0.00	0.193
	ATT Redirects Z	1.34	0.58	0.17	2.30	0.02	
	VTT Redirects Z	0.28	0.64	0.03	0.43	0.67	
	(Constant)	49.13	0.41		119.17	0.00	0.191
	ATT Average Distance Z	-1.24	0.59	-0.15	-2.09	0.04	
	VTT Average Distance Z	-0.58	0.62	-0.07	-0.93	0.35	
	(Constant)	49.13	0.41		119.53	0.00	0.199
	ATT OnTarget Z	1.41	0.58	0.18	2.43	0.02	
	VTT OnTarget Z	0.27	0.64	0.03	0.42	0.68	
	(Constant)	49.08	0.41		118.94	0.00	0.183

	ATT IRT Z	1.25	0.57	0.15	2.19	0.03	
	VTT IRT Z	0.46	0.63	0.05	0.74	0.46	
Instruments ALL	(Constant)	48.78	0.45		107.86	0.00	0.319
	ATT Redirects Z	1.84	0.64	0.21	2.88	0.00	
	VTT Redirects Z	1.40	0.71	0.14	1.97	0.05	
	(Constant)	48.72	0.45		108.43	0.00	0.341
	ATT Average Distance Z	-1.97	0.64	-0.21	-3.06	0.00	
	VTT Average Distance Z	-1.78	0.68	-0.18	-2.60	0.01	
	(Constant)	48.77	0.45		108.08	0.00	0.326
	ATT OnTarget Z	1.98	0.64	0.22	3.10	0.00	
	VTT OnTarget Z	1.33	0.71	0.14	1.89	0.06	
	(Constant)	48.72	0.45		107.75	0.00	0.314
	ATT IRT Z	2.01	0.63	0.21	3.21	0.00	
	VTT IRT Z	1.43	0.69	0.14	2.06	0.04	
Navigation AIRCRAFT	(Constant)	50.28	0.64		78.51	0.00	0.107
	ATT Redirects Z	-0.16	0.86	-0.02	-0.19	0.85	
	VTT Redirects Z	1.18	0.95	0.12	1.24	0.22	
	(Constant)	50.27	0.64		78.47	0.00	0.120
	ATT Average Distance Z	0.62	0.87	0.06	0.72	0.47	
	VTT Average Distance Z	-1.51	0.95	-0.14	-1.60	0.11	
	(Constant)	50.27	0.64		78.45	0.00	0.109
	ATT OnTarget Z	-0.22	0.87	-0.02	-0.25	0.80	
	VTT OnTarget Z	1.26	0.97	0.12	1.31	0.19	
	(Constant)	50.27	0.64		78.86	0.00	0.121
	ATT IRT Z	-0.34	0.83	-0.04	-0.41	0.68	
	VTT IRT Z	1.48	0.95	0.14	1.55	0.12	
Formation AIRCRAFT	(Constant)	49.92	0.57		86.95	0.00	0.287
	ATT Redirects Z	1.30	0.77	0.14	1.69	0.09	
	VTT Redirects Z	1.97	0.88	0.18	2.24	0.03	
	(Constant)	49.86	0.57		87.77	0.00	0.320
	ATT Average Distance Z	-1.50	0.79	-0.15	-1.90	0.06	
	VTT Average Distance Z	-2.33	0.87	-0.21	-2.68	0.01	
	(Constant)	49.90	0.57		87.10	0.00	0.294
	ATT OnTarget Z	1.40	0.77	0.15	1.80	0.07	
	VTT OnTarget Z	1.99	0.89	0.18	2.24	0.03	
	(Constant)	49.87	0.57		86.92	0.00	0.281
	ATT IRT Z	1.25	0.75	0.13	1.66	0.10	
	VTT IRT Z	2.16	0.89	0.19	2.44	0.02	
Navy Standard Score (NSS)	(Constant)	49.03	0.53		91.81	0.00	0.287
	ATT Redirects Z	1.97	0.75	0.19	2.61	0.01	
	VTT Redirects Z	1.45	0.84	0.13	1.73	0.08	

(Constant)	48.98	0.53		92.17	0.00	0.305
ATT Average Distance Z	-2.16	0.77	-0.19	-2.81	0.01	
VTT Average Distance Z	-1.80	0.80	-0.15	-2.24	0.03	
(Constant)	49.02	0.53		91.98	0.00	0.293
ATT OnTarget Z	2.12	0.76	0.20	2.80	0.01	
VTT OnTarget Z	1.40	0.83	0.12	1.68	0.09	
(Constant)	48.98	0.53		91.83	0.00	0.277
ATT IRT Z	2.16	0.74	0.19	2.92	0.00	
VTT IRT Z	1.41	0.81	0.12	1.73	0.08	

In sum, the benefits of combining the ATT and VTT scores across PBM subtests were mixed. Clearly, highly reliable composites are obtained. In addition to improving reliability, applicants' perceptions of the selection process might be more positive because they may perceive greater fairness in that one can compensate for lower performance on one of the subtests with higher performance on another. On the other hand, there is little evidence of enhanced validity accruing from the composites. Thus, it appears that total testing time could be reduced with little effect on validity.

In terms of scoring the responses, IRT seems to offer little advantage for routine operational use. It is much more complex and difficult, but does not appear to enhance validity. Its value may lie chiefly in specialized analyses, such as differential item and test functioning, which could be undertaken on a periodic basis. Of the other three scoring methods, Average Distance seems to yield slightly higher validity, but the differences are not so great as to make it an overwhelming favorite. Ease of computation and ease of use should probably be the deciding factors.

LINEAR COMBINATIONS OF THE PBM SCORES FOR PREDICTION OF TRAINING CRITERIA

Analyses described in earlier sections indicated that various subtest scores of the PBM are predictive of future examinee training grades. In particular, DOT Total Correct scores and most of the ATT and VTT subtest and composite scores had validities of .20 to .30 for multiple training criteria. Other subtests scores such as the MTT DLT, DOT Total Time, and EST Scenario Scores predicted more selectively. For example, the MTT DLT score had high validity for Navigation grades. The only subtest that did not show much promise for selection was DLT. As was noted in the DLT section of this report, there appeared to be issues with the ways this particular subtest was scored. However, given that DLT scores from the MTT subtest were useful, the DLT may still be needed for transitioning into more difficult tracking subtests.

Because none of PBM subtests scores correlated particularly highly with currently used ASTB composites, the use of PBM in conjunction with ASTB should enhance the validity of future selection decisions. The choice of a particular set of PBM scores to augment current Navy pilot selection procedures would ultimately depend on a combination of statistical, practical and policy considerations. In this report, we focus primarily on the statistical side of the process and show the extent to which a chosen subset of PBM scores provides incremental validity for predicting student pilot grades over the PFAR composite, which appears to be designed specifically to predict pilot performance. In the analyses presented below, we selected 6 PBM scores: the two standardized average distance composites representing one-dimensional and two-dimensional tracking abilities (ATT Average Distance Z and VTT Average Distance Z), the DOT Total Correct and DOT Total Time scores representing spatial and processing speed abilities, the MTT DLT Total Correct score representing auditory ability, and EST Scenario measuring stituational awareness and stress tolerance. Relationships between these six PBM scores and other predictors and criteria were investigated in a series of correlational and regression analyses using student pilots. Similar analyses could be performed for other subset of PBM scores that the Navy is considering for selection decisions or with SNFOs when larger samples are available.

Table 85 shows correlations between the six PBM scores. Except for the ATT and VTT, all six PBM predictors have only modest correlations with each other, indicating that they were measuring different abilities.

Table 85. Correlations Between the Six PBM Subtest Scores for Student Pilots (N = 309)

	ATT Average Distance Z	VTT Average Distance Z	DOT Total Correct	DOT Total Time	MTT DLT Total Correct	EST Scenario Score
ATT Average Distance Z	1	.592**	-.296**	.207**	-.218**	-.187**
VTT Average Distance Z	.592**	1	-.325**	.247**	-.259**	-.241**
DOT Total Correct	-.296**	-.325**	1	-.172**	.282**	.243**
DOT Total Time	.207**	.247**	-.172**	1	-.140*	-.153**
MTT DLT Total Correct	-.218**	-.259**	.282**	-.140*	1	.199**
EST Scenario Score	-.187**	-.241**	.243**	-.153**	.199**	1

Table 86 shows correlations between the six PBM scores and various ASTB scores. These correlations were in the -0.35 to 0.35 range indicating that PBM scores measured abilities not measured by ASTB subtests.

Table 86. Correlations Between the Six PBM Subtest Scores and Other Predictors for Student Pilots

	N	Mean	SD	ATT Average Distance Z	VTT Average Distance Z	DOT Total Correct	DOT Total Time	MTT DLT Total Correct	EST Scenario Score
aTraining	305	0.26	0.75	-0.071	-0.045	0.049	0.028	0.021	0
Education	297	2.86	0.64	-0.091	-0.01	0.059	0.049	.188**	-0.017
simExperience	310	0.77	0.85	-.398**	-.305**	.162**	-.119*	0.092	0.045
flightHours	303	0.72	1.42	-0.11	-0.042	-0.018	0.017	-0.033	-0.1
ANI_RAW	248	0.61	0.52	-.216**	-.135*	0.068	-0.06	-0.058	-0.052
MST_RAW	248	0.35	0.67	-0.114	-.234**	.153*	-0.123	.211**	.218**
RCT_RAW	248	0.43	0.54	-0.025	-0.088	0.112	-0.122	.212**	0.11
SAT_Post2004	248	0.80	0.64	-.267**	-.277**	.337**	-.184**	0.121	0.085
MCT_Post2004	248	0.56	0.63	-.226**	-.280**	.330**	-.182**	.200**	.250**
AQR_Post2004	248	0.60	0.51	-.299**	-.345**	.321**	-.207**	.198**	.212**
PFAR_Post2004	248	0.72	0.48	-.337**	-.321**	.311**	-.191**	0.108	0.116
FOFAR_Post2004	248	0.68	0.52	-.294**	-.357**	.317**	-.215**	.210**	.182**
OAR_Post2004	248	0.56	0.62	-.212**	-.303**	.311**	-.191**	.247**	.276**

Table 87 shows correlations between six PBM scores, the PFAR ASTB composite, and training criteria. We included PFAR because it has been specifically designed to predict Pilot Training grades, so any new predictor must be evaluated with respect to this weighted ASTB composite. Note that, for many criteria, these six selected PBM scores performed as well as or better than PFAR. If these score are capturing training variance unrelated to cognitive ability, the PBM scores would likely add significant incremental validities.

Table 87. Correlations Between PFAR, Six PBM Subtest Scores, and Training Grades for Student Pilots.

		Students Pilots (SPs)						
Training Block Name	N	PFAR_Post2004	ATT Average Distance Z	VTT Average Distance Z	DOT Total Correct	DOT Total Time	MTT DLT Total Correct	EST Scenario Score
C20	248	.258**	-.118*	-0.082	.152**	-0.11	.135*	.117*
C40	-	-	-	-	-	-	-	-
C41	234	.229**	-.142*	-.168**	.175**	-0.079	0.039	0.073
C42	227	.151*	-.135*	-.189**	.125*	-0.043	0.074	0.035
C43	215	.170*	-0.094	-0.077	0.098	-0.051	0.109	-0.027
C45	209	.281**	-.168**	-0.12	.174**	-0.116	0.039	0.05
C46	196	.223**	-.214**	-0.122	.150*	-0.104	0.017	-0.001
C47	189	.149*	-0.102	-.132*	0.101	0.073	-0.005	-0.041
I20	243	.278**	-.324**	-.317**	.274**	-.142*	.135*	.131*
I21	240	.252**	-.277**	-.310**	.219**	-.159**	.127*	.159**
I22	170	.210**	-.185**	-.145*	.183**	-.209**	0.106	0.118
I23	170	.207**	-.178*	-.162*	0.116	-.172*	.147*	.186**
I24	160	.284**	-.230**	-.207**	.156*	-.219**	0.1	0.095
I25	156	.327**	-.165*	-0.115	.214**	-0.121	0.055	0.032
I40	239	.257**	-.190**	-.226**	.200**	-.137*	.132*	0.099
I41	165	.227**	-.243**	-.197**	.171*	-.180*	0.076	.165*
I42	152	0.143	-.213**	-.155*	0.072	-0.1	0.107	0.129
I43	148	0.128	-0.078	-0.075	0.071	-0.11	0.029	0.067

F40	183	.291**	-.236**	-.270**	.281**	-.147*	0.127	0.076
F42	172	.293**	-.275**	-.281**	.204**	-.223**	0.061	0.06
N40	153	0.032	-0.047	-0.108	0.069	-0.088	0.136	0.027
N41	151	0.003	0.013	-0.063	0.011	0.004	.168*	0.055
Contact_Simulation	248	.258**	-.118*	-0.082	.152**	-0.11	.135*	.117*
Contact_AIRCRAFT	234	.290**	-.200**	-.206**	.211**	-0.085	0.058	0.05
Contact_ALL	248	.294**	-.179**	-.151**	.217**	-0.09	.115*	0.089
Instruments_Simulation	243	.289**	-.296**	-.294**	.287**	-.184**	.145*	.163**
Instruments_AIRCRAFT	239	.262**	-.255**	-.218**	.226**	-.165**	.150**	.151**
Instruments_ALL	243	.303**	-.308**	-.297**	.297**	-.197**	.159**	.176**
Instruments_BASIC	243	.308**	-.315**	-.341**	.291**	-.185**	.158**	.158**
Instruments_RADIO	170	.250**	-.228**	-.182**	.188**	-.212**	0.121	.170*
Instruments_NAVIGATION	160	.304**	-.240**	-.188**	.182*	-.180*	.144*	0.09
Navigation_AIRCRAFT	153	0.02	-0.018	-0.107	0.034	-0.039	.200**	0.057
Formation_AIRCRAFT	183	.322**	-.268**	-.296**	.275**	-.184**	0.117	0.08
Navy Standard Score (NSS)	248	.354**	-.277**	-.264**	.277**	-.158**	.158**	.156**

Finally, in Table 88 we show multiple regression analyses predicting training criteria with either six PBM scores or a PFAR composite (we only used post 2004 scores), or both. We separated Contact and Instrument grades into Simulator and Aircraft components to better determine where PBM scores are most useful. Specifically, Model 1 includes only PBM predictors; Model 2 includes only PFAR; and Model 3 includes all predictors, PBM-based and ASTB-based.

Comparisons of Models 1, 2 and 3 for each criterion indicated that the use of PBM scores in addition to PFAR resulted in sizable gains for multiple R coefficients. Comparison of these models, regardless of the criterion variables studied, showed significant increments in R. Increments as high as 0.23 were found for predicting Navigation grades, but most of the gains were in the 0.07 to 0.15 range. These results showed that, despite restricted samples (all examinees were already pre-selected for training), PBM can improve prediction, and, therefore, should be considered for operational selection. Unstandardized coefficients, similar to the ones reported for Models 1 or 3 could be used to calculate an overall PBM score and to develop cut scores for selection and/or classification purposes.

Table 88. Regression Models for Predicting Student Pilot Training Grades Using a Combination of PBM Predictors and PFAR

Model	Predictors	N	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	R
			B	Std. Error	Beta			
Contact Simulation	(Constant)	306	43.853	2.993		14.653	.000	.227
	ATT Average Distance Z		-.924	.800	-.082	-1.155	.249	
	VTT Average Distance Z		.691	.864	.058	.799	.425	
	DOT Total Correct		.104	.063	.103	1.662	.098	
	DOT Total Time		-.007	.007	-.060	-1.022	.308	
	MTT DLT Total Correct		.118	.081	.087	1.446	.149	
	EST Scenario Score		.660	.482	.081	1.368	.172	
Contact Simulation	(Constant)	247	45.055	1.086		41.495	.000	.258
	PFAR_Post2004		5.278	1.260	.258	4.191	.000	
Contact Simulation	(Constant)	244	39.798	3.294		12.083	.000	.321
	ATT Average Distance Z		-.657	.905	-.059	-.726	.469	
	VTT Average Distance Z		.852	1.003	.071	.849	.397	
	DOT Total Correct		.090	.071	.089	1.275	.203	
	DOT Total Time		-.005	.008	-.041	-.632	.528	
	MTT DLT Total Correct		.171	.089	.128	1.920	.056	
	EST Scenario Score		.444	.531	.054	.836	.404	
	PFAR_Post2004		4.054	1.372	.200	2.954	.003	
Contact AIRCRAFT	(Constant)	288	45.974	2.361		19.470	.000	.267
	ATT Average Distance Z		-.926	.631	-.107	-1.467	.144	
	VTT Average Distance Z		-.884	.705	-.095	-1.255	.211	
	DOT Total Correct		.123	.050	.156	2.437	.015	
	DOT Total Time		-.002	.005	-.019	-.319	.750	
	MTT DLT Total Correct		-.035	.066	-.032	-.526	.599	
	EST Scenario Score		-.163	.385	-.026	-.424	.672	
Contact AIRCRAFT	(Constant)	233	45.535	.847		53.787	.000	.290

	PFAR_Post2004		4.508	.975	.290	4.623	.000	
Contact AIRCRAFT	(Constant)	230	42.972	2.589		16.596	.000	.335
	ATT Average Distance Z		-.492	.710	-.058	-.693	.489	
	VTT Average Distance Z		-.098	.794	-.011	-.124	.901	
	DOT Total Correct		.127	.056	.163	2.254	.025	
	DOT Total Time		-.003	.006	-.031	-.468	.640	
	MTT DLT Total Correct		-.010	.073	-.009	-.137	.891	
	EST Scenario Score		.028	.421	.004	.066	.947	
	PFAR_Post2004		3.026	1.076	.197	2.813	.005	
Contact ALL	(Constant)	306	44.220	2.225		19.871	.000	.258
	ATT Average Distance Z		-.962	.595	-.113	-1.618	.107	
	VTT Average Distance Z		-.070	.643	-.008	-.109	.913	
	DOT Total Correct		.123	.047	.162	2.630	.009	
	DOT Total Time		-.002	.005	-.020	-.344	.731	
	MTT DLT Total Correct		.041	.061	.040	.677	.499	
	EST Scenario Score		.163	.359	.027	.456	.649	
Contact ALL	(Constant)	247	45.669	.810		56.382	.000	.294
	PFAR_Post2004		4.534	.940	.294	4.825	.000	
Contact ALL	(Constant)	244	40.779	2.450		16.646	.000	.354
	ATT Average Distance Z		-.541	.673	-.064	-.803	.423	
	VTT Average Distance Z		.497	.746	.055	.666	.506	
	DOT Total Correct		.121	.053	.158	2.297	.023	
	DOT Total Time		-.002	.006	-.018	-.282	.778	
	MTT DLT Total Correct		.085	.066	.085	1.287	.199	
	EST Scenario Score		.205	.395	.033	.520	.604	
	PFAR_Post2004		3.286	1.021	.216	3.219	.001	
Instruments Simulation	(Constant)	299	44.440	2.614		16.999	.000	.388
	ATT Average Distance Z		-1.597	.698	-.155	-2.288	.023	
	VTT Average Distance Z		-1.223	.763	-.111	-1.603	.110	
	DOT Total Correct		.154	.055	.166	2.778	.006	

	DOT Total Time		-.009	.006	-.083	-1.481	.140	
	MTT DLT Total Correct		.021	.072	.017	.290	.772	
	EST Scenario Score		.376	.425	.050	.885	.377	
Instruments Simulation	(Constant)	242	44.937	.997		45.056	.000	.289
	PFAR_Post2004		5.407	1.152	.289	4.693	.000	
Instruments Simulation	(Constant)	239	40.894	2.916		14.024	.000	.424
	ATT Average Distance Z		-1.533	.804	-.149	-1.906	.058	
	VTT Average Distance Z		-.202	.899	-.018	-.224	.823	
	DOT Total Correct		.149	.063	.160	2.371	.019	
	DOT Total Time		-.009	.007	-.080	-1.290	.198	
	MTT DLT Total Correct		.098	.080	.079	1.222	.223	
	EST Scenario Score		.412	.472	.054	.873	.383	
	PFAR_Post2004	3	2.641	1.223	.142	2.160	.032	
Instruments AIRCRAFT	(Constant)	294	44.194	2.732		16.178	.000	.330
	ATT Average Distance Z		-1.638	.731	-.157	-2.242	.026	
	VTT Average Distance Z		-.363	.798	-.033	-.455	.650	
	DOT Total Correct		.126	.059	.133	2.155	.032	
	DOT Total Time		-.010	.006	-.087	-1.508	.133	
	MTT DLT Total Correct		.055	.075	.043	.725	.469	
	EST Scenario Score		.504	.447	.066	1.128	.260	
Instruments AIRCRAFT	(Constant)	238	45.056	1.032		43.646	.000	.262
	PFAR_Post2004		4.961	1.187	.262	4.178	.000	
Instruments AIRCRAFT	(Constant)	235	42.788	3.077		13.906	.000	.372
	ATT Average Distance Z		-1.539	.851	-.147	-1.809	.072	
	VTT Average Distance Z		.033	.948	.003	.035	.972	
	DOT Total Correct		.112	.067	.117	1.664	.097	
	DOT Total Time		-.010	.007	-.086	-1.344	.180	
	MTT DLT Total Correct		.044	.085	.035	.521	.603	
	EST Scenario Score		.639	.501	.082	1.273	.204	
	PFAR_Post2004		2.654	1.294	.140	2.051	.041	

Instruments ALL	(Constant)	299	44.268	2.396		18.473	.000	.405
	ATT Average Distance Z		-1.574	.640	-.165	-2.461	.014	
	VTT Average Distance Z		-1.001	.699	-.099	-1.431	.154	
	DOT Total Correct		.148	.051	.173	2.927	.004	
	DOT Total Time		-.009	.006	-.093	-1.664	.097	
	MTT DLT Total Correct		.028	.066	.025	.433	.665	
	EST Scenario Score		.427	.389	.062	1.096	.274	
Instruments ALL	(Constant)	242	44.968	.917		49.062	.000	.303
	PFAR_Post2004		5.226	1.059	.303	4.935	.000	
Instruments ALL	(Constant)	239	41.375	2.672		15.484	.000	.445
	ATT Average Distance Z		-1.486	.737	-.156	-2.016	.045	
	VTT Average Distance Z		-.182	.824	-.018	-.221	.825	
	DOT Total Correct		.145	.058	.168	2.513	.013	
	DOT Total Time		-.009	.006	-.093	-1.509	.133	
	MTT DLT Total Correct		.077	.073	.067	1.051	.294	
	EST Scenario Score		.503	.433	.072	1.161	.247	
	PFAR_Post2004		2.555	1.121	.149	2.280	.024	
Navigation AIRCRAFT	(Constant)	181	46.056	3.910		11.778	.000	.219
	ATT Average Distance Z		.811	.871	.084	.931	.353	
	VTT Average Distance Z		-1.144	1.014	-.109	-1.128	.261	
	DOT Total Correct		-.041	.079	-.042	-.518	.605	
	DOT Total Time		.001	.008	.013	.162	.871	
	MTT DLT Total Correct		.262	.108	.197	2.437	.016	
	EST Scenario Score		.074	.550	.010	.135	.893	
Navigation AIRCRAFT	(Constant)	152	49.904	1.283		38.906	.000	.020
	PFAR_Post2004		.347	1.417	.020	.245	.807	
Navigation AIRCRAFT	(Constant)	151	45.060	4.350		10.358	.000	.251
	ATT Average Distance Z		.475	1.041	.047	.457	.649	
	VTT Average Distance Z		-1.085	1.173	-.103	-.925	.356	

	DOT Total Correct		-.035	.091	-.035	-.383	.702	
	DOT Total Time		.002	.010	.013	.157	.876	
	MTT DLT Total Correct		.320	.121	.235	2.641	.009	
	EST Scenario Score		-.022	.638	-.003	-.035	.972	
	PFAR_Post2004		-.645	1.603	-.037	-.402	.688	
Formation AIRCRAFT	(Constant)	223	47.033	3.110		15.123	.000	.373
	ATT Average Distance Z		-1.178	.789	-.117	-1.493	.137	
	VTT Average Distance Z		-1.531	.917	-.138	-1.669	.097	
	DOT Total Correct		.174	.067	.186	2.595	.010	
	DOT Total Time		-.012	.008	-.105	-1.593	.113	
	MTT DLT Total Correct		-.025	.089	-.019	-.282	.778	
	EST Scenario Score		-.160	.493	-.022	-.324	.746	
Formation AIRCRAFT	(Constant)	182	46.033	1.032		44.594	.000	.322
	PFAR_Post2004		5.206	1.136	.322	4.584	.000	
Formation AIRCRAFT	(Constant)	181	46.405	3.177		14.608	.000	.406
	ATT Average Distance Z		-.949	.854	-.101	-1.112	.268	
	VTT Average Distance Z		-.677	.962	-.067	-.704	.483	
	DOT Total Correct		.105	.071	.122	1.477	.141	
	DOT Total Time		-.012	.008	-.116	-1.583	.115	
	MTT DLT Total Correct		.001	.094	.001	.012	.990	
	EST Scenario Score		-.004	.519	-.001	-.007	.994	
	PFAR_Post2004		3.031	1.281	.188	2.367	.019	
Navy Standard Score (NSS)	(Constant)	306	42.391	2.844		14.906	.000	.372
	ATT Average Distance Z		-1.693	.760	-.150	-2.228	.027	
	VTT Average Distance Z		-.914	.821	-.077	-1.112	.267	
	DOT Total Correct		.180	.060	.178	3.008	.003	
	DOT Total Time		-.007	.007	-.058	-1.040	.299	
	MTT DLT Total Correct		.053	.077	.039	.684	.494	
	EST Scenario Score		.495	.458	.061	1.080	.281	

Navy Standard Score (NSS)	(Constant)	247	43.767	1.038		42.175	.000	.354
	PFAR_Post2004		7.142	1.204	.354	5.933	.000	
Navy Standard Score (NSS)	(Constant)	244	37.575	3.065		12.259	.000	.460
	ATT Average Distance Z		-1.332	.842	-.120	-1.582	.115	
	VTT Average Distance Z		.128	.934	.011	.137	.891	
	DOT Total Correct		.165	.066	.164	2.510	.013	
	DOT Total Time		-.007	.007	-.056	-.937	.350	
	MTT DLT Total Correct		.134	.083	.101	1.609	.109	
	EST Scenario Score		.604	.494	.074	1.222	.223	
	PFAR_Post2004		4.614	1.277	.230	3.613	.000	

SUMMARY AND CONCLUSION

PBM subtest level scores appeared to provide valuable information for predicting training outcomes and resulted in substantial validities with various training criteria, especially those concerned with the actual operation of aircraft. The IRT analyses indicated that the three parameter logistic model (3PLM) and the Samejima's graded response model (SGRM) provided good fit to dichotomously and polytomously scored item-level data, paving the way for future research involving differential item and test functioning. However, in terms of validities, subtest level CTT-based multiple regression composites seemed to perform best and are thus recommended for operational decision making. The inclusion of a PBM composite similar to a six variable set discussed in the previous chapter would be beneficial for selection of student pilots. The CTT based PBM composite that best predicted pilot primary flight school performance in the current sample was composed of the following PBM sub-scores: Airplane Tracking Task Average Distance Z-score, the Vertical Tracking Task Average Distance Z-score, the Directional Orientation Test Total Correct, the Directional Orientation Test Total Time, the Multi Tracking Test Dichotic Listening Tests Total Correct, and the Emergency Scenarios Test Scenario Score. Sample sizes for SNFOs were too small to derive PBM-based composites for that group, but the magnitude of observed correlations between PBM component scores and training criteria were similar to those observed for SPs. PBM scores had only moderate correlations with ASTB scores indicating that abilities measured by PBM are not currently being captured by ASTB.

The marked increase in incremental validity that results from the addition of PBM composites to the ASTB suggests that the addition of the PBM to Naval Aviation selection will significantly reduce attrition from the Naval Aviation training pipeline and save Naval Aviation millions in training costs. Because our analysis did not include a hold-out sample, the predictive validity of this composite score should be confirmed using a new sample of Naval Aviation students.

References

Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F.M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397-472). Reading, MA: Addison-Wesley.

Carretta, T. R (2005). *Development and validation of the Test of Basic Aviation Skills (TBAS).* (AFRL-HE-WP-TR-2005-0172). Wright-Patterson AFB, OH: Air Force Research Laboratory, Human Effectiveness Directorate.

Carretta, T. R., & Ree, M. J. (1993). Pilot candidate selection method (PCSM): What makes it work? (AL-TP-1992-0063). Brooks AFB, TX: Manpower and Personnel Research Division, Human Resources Directorate, Armstrong Laboratory.

Drasgow, F., Levine M. V., Tsien, S., Williams, B. A., & Mead, A. D. (1995). Fitting polytomous item response theory models to multiple-choice tests. *Applied Psychological Measurement,19*, 143 – 165.

Drasgow, F., & Parsons, C. K. (1983). Application of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement,7,* 189 – 199.

Imhoff, D. F., & Levine, J. M. (1981). *Perceptual-motor and cognitive performance task battery for pilot selection* (Tech. Rep. No. AFHRL-TR-80-27). Brooks Air Force Base, TX: Air Force Human Resources Laboratory, Manpower and Personnel Division.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Melton, A. W. (Ed.) (1947). *Apparatus tests*. Air Force Aviation Psychology Program Research Report No. 4. Washington, DC: U.S. Government Printing Office.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement*, *34*, 100-114.

Stark, S. (2007). *MODFIT-Z 2.0: An Excel VBA program for examining item and person fit and scoring with dichotomous and polytomous IRT models.* Drasgow Consulting Group.

Thissen, D. (1991). *MULTILOG user's guide (Version 6)* [Computer program]. Mooresville, IN: Scientific Software Inc.

Unpingco, V., Hom, I., & Rafacz, B. (1997). Development of a system for nationwide implementation. In W. A. Sands, B. K. Waters, & J. R. McBride (Eds.), *Computerized adaptive testing: From inquiry to operation* (pp. 209-218). Washington, DC: American Psychological Association.

Williams, H. P., Albert, A. O., & Blower, D. J. (November, 1999). Selection of officers for U.S. Naval Aviation Training. Paper presented at the RTO HFM Workshop on "Officer Selection", Monterey, CA.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (1996). BILOG-MG. [Computer program]. Chicago, IL: Scientific Software Inc.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (DD MM YY)
29 11 11

2. REPORT TYPE
Technical Report

3. DATES COVERED (from – to)
Oct 2007 – Oct 2009

4. TITLE
Development of Scoring Procedures for the Performance Based Measurement (PBM) Test: Psychometric and Criterion Validity Investigation

5a. Contract Number:
5b. Grant Number:
5c. Program Element Number:
5d. Project Number:
5e. Task Number:
5f. Work Unit Number: 70804

6. AUTHORS
Phillips, J.B., Chernyshenko, O.S., Stark, S., Drasgow, F., & Phillips, H.L.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Medical Research Unit – Dayton
2624 Q St., Bld. 851, Area B
WPAFB, OH 45433-7955

8. PERFORMING ORGANIZATION REPORT NUMBER
Report No. NAMRU-D-12-10

8. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)
Office of Naval Research, Code 34
One Liberty Center
875 N Randolf St., Ste. 1425
Arlington, VA 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR Code 34

11. SPONSOR/MONITOR'S REPORT NUMBER(s)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
N/A

14. ABSTRACT

This report describes findings regarding the scoring of the PBM Test and the relationships of various classical test theory (CTT) and item response theory (IRT) based subtest scores and composites with performance criteria for Navy and Marine Corps student pilots and flight officers. Overall, the IRT analyses indicated that the three parameter logistic model (3PLM) and Samejima's graded response model (SGRM) provided good fit to dichotomously and polytomously scored item-level data, respectively, for six of the seven PBM subtests. These analyses set the stage for future research involving differential item and test functioning. It was also found that PBM component scores based on examinee responses, reaction time, or tracking information often yielded criterion related validities in the range of .20 - .35. IRT-based scores did not outperform CTT-based scores, therefore we recommend using subtest level CTT-based multiple regression composites for operational decision making. These results suggest that the addition of the PBM to ASTB for Naval Aviation selection will significantly reduce attrition from the Naval Aviation training pipeline, thereby saving Naval Aviation millions in training costs. The following sections of this report describe the PBM subtests, the demographics of the examinee samples, and the results of the psychometric and statistical analyses showing that the PBM subtests are valid predictors of a wide variety of training criteria.

15. SUBJECT TERMS
performance based measurement test, aviation selection, psychomotor ability, spatial apperception

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	18a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Commanding Officer
UNCL	UNCL	UNCL	UNCL	146	18b. TELEPHONE NUMBER (INCLUDING AREA CODE) COMM/DSN: 937-938-3872 (DSN: 798)